



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification 6 : C12N 15/54, 15/82, 15/11, 9/12, 5/10, C12Q 1/68</p>	<p>A1</p>	<p>(43) International Publication Number: WO 99/55879 (43) International Publication Date: 4 November 1999 (04.11.99)</p>
<p>(21) International Application Number: PCT/US99/08790 (22) International Filing Date: 22 April 1999 (22.04.99)</p>		<p>(81) Designated States: AE, AL, AU, BA, BB, BG, BR, CA, CN, CU, CZ, EE, GD, GE, HR, HU, ID, IL, IN, IS, JP, KR, LC, LK, LR, LT, LV, MG, MK, MN, MX, NO, NZ, PL, RO, SG, SI, SK, SL, TR, TT, UA, US, UZ, VN, YU, ZA, ARIPO patent (GH, GM, KE, LS, MW, SD, SI, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</p>
<p>(30) Priority Data: 60/082,960 24 April 1998 (24.04.98) US</p>		
<p>(71) Applicant (for all designated States except US): E.I. DU PONT DE NEMOURS AND COMPANY [US/US]; 1007 Market Street, Wilmington, DE 19898 (US).</p>		
<p>(72) Inventors; and</p>	<p>Published</p>	
<p>(75) Inventors/Applicants (for US only): CAHOON, Rebecca, E. [US/US]; 231 18th Street, Wilmington, DE 19806 (US); CARLSON, Thomas, J. [US/US]; 2316 Orchard Road, Wilmington, DE 19810 (US); HITZ, William, D. [US/US]; 404 Hillside Road, Wilmington, DE 19807 (US); PEARLSTEIN, Richard, W. [US/US]; 12 Canoe Court, Newark, DE 19702 (US).</p>	<p>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</p>	
<p>(74) Agent: MAJARIAN, William, R.; E.I. du Pont de Nemours and Company, Legal Patent Records Center, 1007 Market Street, Wilmington, DE 19898 (US).</p>		
<p>(54) Title: PHYTIC ACID BIOSYNTHETIC ENZYMES</p>		

(57) Abstract

This invention relates to an isolated nucleic acid fragment encoding a phytic acid biosynthetic enzyme. The invention also relates to the construction of a chimeric gene encoding all or a portion of the phytic acid biosynthetic enzyme, in sense or antisense orientation, wherein expression of the chimeric gene results in production of altered levels of the phytic acid biosynthetic enzyme in a transformed host cell.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Cote d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

TITLE

PHYTIC ACID BIOSYNTHETIC ENZYMES

This application claims the benefit of U.S. Provisional Application No. 60/082,960, filed April 24, 1998.

FIELD OF THE INVENTION

This invention is in the field of plant molecular biology. More specifically, this invention pertains to nucleic acid fragments encoding phytic acid biosynthetic enzymes in plants and seeds.

BACKGROUND OF THE INVENTION

Myo-inositol 1,2,3,4,5,6-hexaphosphate, commonly known as phytic acid, is an abundant molecule in many plant seeds and vegetative tissue such as roots and tubers (Hartland and Oberlaas, (1986) *J. Assoc. Off. Anal. Chem.* 69:667-670). Phytic acid exists primarily as mixture of potassium, calcium, iron, zinc and magnesium phytate salts (Pernollet J. C. (1978) *Phytochemistry* 17:1473-1480).

In corn (*Zea mays* L.), 90% of the phytate is deposited in protein bodies localized in the germ whereas in legume crops 90% of the phytate is localized in the endosperm and cotyledons. Up to 80% of phytate is in the aleurone layer of wheat (*Triticum aestivum* Lam.) and rice (*Oryza sativa* L.) (O'Dell B. L. et al. (1972) *J. Agric. Food Chem.* 20:718-721). The presence of phytate phosphorous in such food crops decreases the bioavailability of zinc by forming a very stable insoluble phytate zinc complex, making the zinc unavailable in the intestinal mucosa of mammals (O'Dell, B. L., et al. (1972) *J. Agr. Food Chem.* 20:718-721). Although phytate phosphorous is readily available to ruminants, it is less available to monogastric animals. In addition to being only partially digestible, the presence of phytic acid in food crops leads to excretion of other limiting nutrients such as essential amino acids, calcium and zinc (Mroz, Z. et al. (1994) *J. Animal Sci.* 72:126-132; Fox et al., In Nutritional Toxicology Vol. 3, Academic Press, San Diego (1989) pp. 59-96).

Phytic acid is thought to arise in plants by two pathways. The first pathway uses free *myo*-inositol as the initial substrate, with subsequent phosphorylation by a phosphoinositol kinase. Contribution to the free *myo*-inositol pool is either by recycling from other pathways or by the dephosphorylation of *myo*-inositol-1-phosphate. The alternate pathway uses *myo*-inositol-1-phosphate as the initial substrate, with subsequent phosphorylations catalyzed by phosphoinositol kinase. The committed step for *myo*-inositol-1-phosphate production is the NAD⁺-catalyzed oxidation of carbon 5 of the D-glucose-6-phosphate. This reaction is catalyzed by *myo*-inositol-1-phosphate synthase (Raboy, V. In Inositol Metabolism in Plants (1990) Wiley-Liss, New York, pp. 55-76).

Phytic acid is degraded in plant cells to D-*myo*-inositol 1,2,4,5,6-pentakisphosphate and orthophosphate through the action of phytase. Manipulation of this enzyme activity could lead to a reduction of phytic acid levels in seeds and an increase in inositol trisphosphate and free phosphate, thus making phosphorus more metabolically available to

animals that are fed the seed. Another method to lower phytic acid levels is by inhibiting the activity of myo-inositol-1(or 4)-monophosphatase, which catalyzes the reaction: myo-inositol 1-phosphate + H₂O = myo-inositol + orthophosphate. Manipulation of the activity of this enzyme in developing seeds could decrease phytic acid levels in seeds and increase levels of free phosphate. Lastly, phytic acid levels could also be reduced by inhibiting the activity of inositol trisphosphate kinase. This enzyme catalyzes the reaction: ATP + 1D-myo-inositol 1,3,4-trisphosphate = ADP + 1D-myo-inositol 1,3,4,6-tetrakisphosphate. This reaction is one of the final steps leading to the formation of Myo-Inositol 1,2,3,4,5,6-hexaphosphate (phytic acid). Reduction in the activity of the enzyme in developing seeds would interrupt phytic acid synthesis leaving the phosphate as the more metabolically available inositol trisphosphate and free phosphate.

In the United States, corn accounts for about 80% of the grain fed to all classes of livestock, including poultry, and is usually ground before feeding (Corn: Chemistry and Technology, 1987, American Association of Cereal Chemists, Inc., Edited by Stanley A. Watson and Paul E. Ramstad). A meal with decreased amounts of phytic acid and increased amounts of available phosphate would lead to improved feed efficiency in corn-containing rations, making available certain minerals especially zinc, magnesium, iron and calcium. Indeed, enzymatic treatment of soybean meal-containing rations to partially hydrolyze the phosphate groups from phytic acid improves both phosphate availability and the availability of other limiting nutrients. Also, in the wet milling of corn, phytate in the steepwater tends to precipitate, causing problems in handling, storing and transportation of the steep liquor. (Pen et al. (1993) *Biotechnology* 11:811-814). In light of these factors, it is apparent that corn plants with heritable, substantially reduced levels of phytic acid and increased levels of free phosphorous in their seeds would be desirable. Accordingly, the availability of nucleic acid sequences encoding all or a portion of these enzymes would facilitate studies to better understand carbohydrate metabolism and function in plants, provide genetic tools for the manipulation of these biosynthetic pathways, and provide a means to control carbohydrate transport and distribution in plant cells.

SUMMARY OF THE INVENTION

The instant invention relates to isolated nucleic acid fragments encoding phytic acid biosynthetic enzymes. Specifically, this invention concerns an isolated nucleic acid fragment encoding an inositol 1,3,4-trisphosphate 5/6-kinase. In addition, this invention relates to a nucleic acid fragment that is complementary to the nucleic acid fragment encoding inositol 1,3,4-trisphosphate 5/6-kinase. An additional embodiment of the instant invention pertains to a polypeptide encoding all or a substantial portion of an inositol 1,3,4-trisphosphate 5/6-kinase.

In another embodiment, the instant invention relates to a chimeric gene encoding an inositol 1,3,4-trisphosphate 5/6-kinase, or to a chimeric gene that comprises a nucleic acid fragment that is complementary to a nucleic acid fragment encoding an inositol

1,3,4-triphosphate 5/6-kinase, operably linked to suitable regulatory sequences, wherein expression of the chimeric gene results in production of levels of the encoded protein in a transformed host cell that is altered (i.e., increased or decreased) from the level produced in an untransformed host cell.

- 5 In a further embodiment, the instant invention concerns a transformed host cell comprising in its genome a chimeric gene encoding an inositol 1,3,4-triphosphate 5/6-kinase, operably linked to suitable regulatory sequences. Expression of the chimeric gene results in production of altered levels of the encoded protein in the transformed host cell. The transformed host cell can be of eukaryotic or prokaryotic origin, and include cells
10 derived from higher plants and microorganisms. The invention also includes transformed plants that arise from transformed host cells of higher plants, and seeds derived from such transformed plants.

- An additional embodiment of the instant invention concerns a method of altering the level of expression of an inositol 1,3,4-triphosphate 5/6-kinase in a transformed host cell
15 comprising: a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding an inositol 1,3,4-triphosphate 5/6-kinase; and b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the chimeric gene results in production of altered levels of inositol 1,3,4-triphosphate 5/6-kinase in the transformed host cell.

- 20 An addition embodiment of the instant invention concerns a method for obtaining a nucleic acid fragment encoding all or a substantial portion of an amino acid sequence encoding an inositol 1,3,4-triphosphate 5/6-kinase.

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE DESCRIPTIONS

- 25 The invention can be more fully understood from the following detailed description and the accompanying drawings and Sequence Listing which form a part of this application.

- Figure 1 shows a comparison of the amino acid sequence set forth in SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30 and 32 with the *Arabidopsis thaliana* inositol 1,3,4-triphosphate 5/6-kinase amino acid sequences set forth in NCBI Identifier
30 No. gi 3396079 (SEQ ID NO:33) and NCBI Identifier No. gi 3660465 (SEQ ID NO:34). Alignments were performed using the Clustal algorithm.

- The following sequence descriptions and sequence listings attached hereto comply with the rules governing nucleotide and/or amino acid sequence disclosures in patent applications as set forth in 37 C.F.R. §1.821-1.825.

- 35 SEQ ID NO:1 is the nucleotide sequence comprising a contig assembled from the cDNA inserts in clones cca.pk0022.e6, cpe1c.pk001.h8, cr1n.pk0188.g8, p0005.cbmf77r and p0090.cspg46r encoding a corn inositol 1,3,4-triphosphate 5/6-kinase.

SEQ ID NO:2 is the deduced amino acid sequence of an inositol 1,3,4-triphosphate 5/6-kinase derived from the nucleotide sequence of SEQ ID NO:1.

SEQ ID NO:3 is the nucleotide sequence comprising a portion of the cDNA insert in clone dms2c.pk003.m14 encoding a portion of an african daisy inositol 1,3,4-triphosphate 5/6-kinase.

5 SEQ ID NO:4 is the deduced amino acid sequence of a portion of an inositol 1,3,4-triphosphate 5/6-kinase derived from the nucleotide sequence of SEQ ID NO:3.

SEQ ID NO:5 is the nucleotide sequence comprising a portion of the cDNA insert in clone ncs.pk0019.a6 encoding a portion of a *Catalpa* inositol 1,3,4-triphosphate 5/6-kinase.

10 SEQ ID NO:6 is the deduced amino acid sequence of a portion of an inositol 1,3,4-triphosphate 5/6-kinase derived from the nucleotide sequence of SEQ ID NO:5.

SEQ ID NO:7 is the nucleotide sequence comprising a contig assembled from the cDNA inserts in clones p0125.czaaj15r, p0125.czabg28r, p0125.czabp82r and p0041.ertcl17r encoding a corn inositol 1,3,4-triphosphate 5/6-kinase.

SEQ ID NO:8 is the deduced amino acid sequence of an inositol 1,3,4-triphosphate 5/6-kinase derived from the nucleotide sequence of SEQ ID NO:8.

15 SEQ ID NO:9 is the nucleotide sequence comprising a contig assembled from the cDNA inserts in clones rr1.pk0052.fl and rl0n.pk0015.b2 encoding a portion of a rice inositol 1,3,4-triphosphate 5/6-kinase.

SEQ ID NO:10 is the deduced amino acid sequence of a portion of an inositol 1,3,4-triphosphate 5/6-kinase derived from the nucleotide sequence of SEQ ID NO:9.

20 SEQ ID NO:11 is the nucleotide sequence comprising the entire cDNA insert in clone rlr12.pk0012.c11 encoding a rice inositol 1,3,4-triphosphate 5/6-kinase.

SEQ ID NO:12 is the deduced amino acid sequence of an inositol 1,3,4-triphosphate 5/6-kinase derived from the nucleotide sequence of SEQ ID NO:11.

25 SEQ ID NO:13 is the nucleotide sequence comprising a portion of the cDNA insert in clone rr1.pk0061.c5 encoding a portion of a rice inositol 1,3,4-triphosphate 5/6-kinase.

SEQ ID NO:14 is the deduced amino acid sequence of a portion of an inositol 1,3,4-triphosphate 5/6-kinase derived from the nucleotide sequence of SEQ ID NO:13.

30 SEQ ID NO:15 is the nucleotide sequence comprising the entire cDNA insert in clone sfl1.pk0091.c9 encoding a soybean inositol 1,3,4-triphosphate 5/6-kinase.

SEQ ID NO:16 is the deduced amino acid sequence of an inositol 1,3,4-triphosphate 5/6-kinase derived from the nucleotide sequence of SEQ ID NO:15.

SEQ ID NO:17 is the nucleotide sequence comprising the entire cDNA insert in clone sgs3n.pk001.b5 encoding a soybean inositol 1,3,4-triphosphate 5/6-kinase.

35 SEQ ID NO:18 is the deduced amino acid sequence of an inositol 1,3,4-triphosphate 5/6-kinase derived from the nucleotide sequence of SEQ ID NO:17.

SEQ ID NO:19 is the nucleotide sequence comprising a portion of the cDNA insert in clone sl1.pk0026.a8 encoding a portion of a soybean inositol 1,3,4-triphosphate 5/6-kinase.

SEQ ID NO:20 is the deduced amino acid sequence of a portion of an inositol 1,3,4-triphosphate 5/6-kinase derived from the nucleotide sequence of SEQ ID NO:19.

SEQ ID NO:21 is the nucleotide sequence comprising a portion of the cDNA insert in clone sls2c.pk013.j24 encoding a portion of a soybean inositol 1,3,4-triphosphate 5/6-kinase.

SEQ ID NO:22 is the deduced amino acid sequence of a portion of an inositol 1,3,4-triphosphate 5/6-kinase derived from the nucleotide sequence of SEQ ID NO:21.

5 SEQ ID NO:23 is the nucleotide sequence comprising a portion of the cDNA insert in clone wdk4c.pk005.a15(5') encoding a portion of a wheat inositol 1,3,4-triphosphate 5/6-kinase.

SEQ ID NO:24 is the deduced amino acid sequence of a portion of an inositol 1,3,4-triphosphate 5/6-kinase derived from the nucleotide sequence of SEQ ID NO:23.

10 SEQ ID NO:25 is the nucleotide sequence comprising a portion of the cDNA insert in clone wdk4c.pk005.a15(3') encoding a portion of a wheat inositol 1,3,4-triphosphate 5/6-kinase.

SEQ ID NO:26 is the deduced amino acid sequence of a portion of an inositol 1,3,4-triphosphate 5/6-kinase derived from the nucleotide sequence of SEQ ID NO:25.

15 SEQ ID NO:27 is the nucleotide sequence comprising a portion of the cDNA insert in clone wr1.pk0137.c5(5') encoding a portion of a wheat inositol 1,3,4-triphosphate 5/6-kinase.

SEQ ID NO:28 is the deduced amino acid sequence of a portion of an inositol 1,3,4-triphosphate 5/6-kinase derived from the nucleotide sequence of SEQ ID NO:27.

20 SEQ ID NO:29 is the nucleotide sequence comprising a portion of the cDNA insert in clone wr1.pk0137.c5(3') encoding a portion of a wheat inositol 1,3,4-triphosphate 5/6-kinase.

SEQ ID NO:30 is the deduced amino acid sequence of a portion of an inositol 1,3,4-triphosphate 5/6-kinase derived from the nucleotide sequence of SEQ ID NO:30.

25 SEQ ID NO:31 is the nucleotide sequence comprising a portion of the cDNA insert in clone wr1.pk0150.e10 encoding a portion of a wheat inositol 1,3,4-triphosphate 5/6-kinase.

SEQ ID NO:32 is the deduced amino acid sequence of a portion of an inositol 1,3,4-triphosphate 5/6-kinase derived from the nucleotide sequence of SEQ ID NO:31.

30 The Sequence Listing contains the one letter code for nucleotide sequence characters and the three letter codes for amino acids as defined in conformity with the IUPAC-IUBMB standards described in *Nucleic Acids Research* 13:3021-3030 (1985) and in the *Biochemical Journal* 219 (No. 2):345-373 (1984) which are herein incorporated by reference. The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

35 DETAILED DESCRIPTION OF THE INVENTION

In the context of this disclosure, a number of terms shall be utilized. As used herein, an "isolated nucleic acid fragment" is a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. An isolated nucleic acid fragment in the form of a polymer of DNA may be comprised of one or

more segments of cDNA, genomic DNA or synthetic DNA. As used herein, "contig" refers to an assemblage of overlapping nucleic acid sequences to form one contiguous nucleotide sequence. For example, several DNA sequences can be compared and aligned to identify common or overlapping regions. The individual sequences can then be assembled into a single contiguous nucleotide sequence.

As used herein, "substantially similar" refers to nucleic acid fragments wherein changes in one or more nucleotide bases results in substitution of one or more amino acids, but do not affect the functional properties of the protein encoded by the DNA sequence.

"Substantially similar" also refers to nucleic acid fragments wherein changes in one or more nucleotide bases does not affect the ability of the nucleic acid fragment to mediate alteration of gene expression by antisense or co-suppression technology. "Substantially similar" also refers to modifications of the nucleic acid fragments of the instant invention such as deletion or insertion of one or more nucleotides that do not substantially affect the functional properties of the resulting transcript vis-à-vis the ability to mediate alteration of gene expression by antisense or co-suppression technology or alteration of the functional properties of the resulting protein molecule. It is therefore understood that the invention encompasses more than the specific exemplary sequences.

For example, it is well known in the art that antisense suppression and co-suppression of gene expression may be accomplished using nucleic acid fragments representing less than the entire coding region of a gene, and by nucleic acid fragments that do not share 100% sequence identity with the gene to be suppressed. Moreover, alterations in a gene which result in the production of a chemically equivalent amino acid at a given site, but do not effect the functional properties of the encoded protein, are well known in the art. Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to produce a functionally equivalent product. Nucleotide changes which result in alteration of the N-terminal and C-terminal portions of the protein molecule would also not be expected to alter the activity of the protein. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products.

Moreover, substantially similar nucleic acid fragments may also be characterized by their ability to hybridize, under stringent conditions (0.1X SSC, 0.1% SDS, 65°C), with the nucleic acid fragments disclosed herein.

Substantially similar nucleic acid fragments of the instant invention may also be characterized by the percent similarity of the amino acid sequences that they encode to the amino acid sequences disclosed herein, as determined by algorithms commonly employed by

- those skilled in this art. Preferred are those nucleic acid fragments whose nucleotide sequences encode amino acid sequences that are 80% similar to the amino acid sequences reported herein. More preferred nucleic acid fragments encode amino acid sequences that are 90% similar to the amino acid sequences reported herein. Most preferred are nucleic acid fragments that encode amino acid sequences that are 95% similar to the amino acid sequences reported herein. Sequence alignments and percent similarity calculations were performed using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences was performed using the Clustal method of alignment (Higgins, D. G. and Sharp, P. M. (1989) *CABIOS* 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10) (hereafter, Clustal algorithm). Default parameters for pairwise alignments using the Clustal method were KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

- A "substantial portion" of an amino acid or nucleotide sequence comprises enough of the amino acid sequence of a polypeptide or the nucleotide sequence of a gene to afford putative identification of that polypeptide or gene, either by manual evaluation of the sequence by one skilled in the art. or by computer-automated sequence comparison and identification using algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al., (1993) *J. Mol. Biol.* 215:403-410; see also www.ncbi.nlm.nih.gov/BLAST/). In general, a sequence of ten or more contiguous amino acids or thirty or more nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene specific oligonucleotide probes comprising 20-30 contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12-15 bases may be used as amplification primers in PCR in order to obtain a particular nucleic acid fragment comprising the primers. Accordingly, a "substantial portion" of a nucleotide sequence comprises enough of the sequence to afford specific identification and/or isolation of a nucleic acid fragment comprising the sequence. The instant specification teaches partial or complete amino acid and nucleotide sequences encoding one or more particular plant proteins. The skilled artisan, having the benefit of the sequences as reported herein, may now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this art. Accordingly, the instant invention comprises the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above.

"Codon degeneracy" refers to divergence in the genetic code permitting variation of the nucleotide sequence without effecting the amino acid sequence of an encoded polypeptide. Accordingly, the instant invention relates to any nucleic acid fragment that

encodes all or a substantial portion of the amino acid sequence encoding the inositol 1,3,4-triphosphate 5/6-kinase proteins as set forth in SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30 and 32. The skilled artisan is well aware of the "codon-bias" exhibited by a specific host cell in usage of nucleotide codons to specify a given amino acid.

- 5 Therefore, when synthesizing a gene for improved expression in a host cell, it is desirable to design the gene such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

"Synthetic genes" can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form gene segments which are then enzymatically assembled to construct the entire gene. "Chemically synthesized", as related to a sequence of DNA, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of DNA may be accomplished using well established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the genes can be tailored for optimal gene expression based on optimization of nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

- 20 "Gene" refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. "Native gene" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene" refers any gene that is not a native gene, comprising regulatory and coding sequences that are not found together in nature.
- 25 Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. "Endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign" gene refers to a gene not normally found in the host organism, but
- 30 that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A "transgene" is a gene that has been introduced into the genome by a transformation procedure.

"Coding sequence" refers to a DNA sequence that codes for a specific amino acid sequence. "Regulatory sequences" refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, and polyadenylation recognition sequences.

"Promoter" refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. The promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an "enhancer" is a DNA sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. Promoters which cause a gene to be expressed in most cell types at most times are commonly referred to as "constitutive promoters". New promoters of various types useful in plant cells are constantly being discovered; numerous examples may be found in the compilation by Okamuro and Goldberg, (1989) *Biochemistry of Plants* 15:1-82. It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths may have identical promoter activity.

The "translation leader sequence" refers to a DNA sequence located between the promoter sequence of a gene and the coding sequence. The translation leader sequence is present in the fully processed mRNA upstream of the translation start sequence. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency. Examples of translation leader sequences have been described (Turner, R. and Foster, G. D. (1995) *Molecular Biotechnology* 3:225).

The "3' non-coding sequences" refer to DNA sequences located downstream of a coding sequence and include polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht et al., (1989) *Plant Cell* 1:671-680.

"RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA (mRNA)" refers to the RNA that is without introns and that can be translated into protein by the cell. "cDNA" refers to a double-stranded DNA that is complementary to and derived from mRNA. "Sense" RNA refers to RNA transcript that includes the mRNA and so can be translated into protein by the cell. "Antisense RNA" refers to a RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene (U.S.

Patent No. 5,107,065, incorporated herein by reference). The complementarity of an antisense RNA may be with any part of the specific gene transcript, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. "Functional RNA" refers to sense RNA, antisense RNA, ribozyme RNA, or other RNA that may not be translated but yet has an effect on cellular processes.

The term "operably linked" refers to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term "expression", as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. Expression may also refer to translation of mRNA into a polypeptide.

"Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of the target protein. "Overexpression" refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms. "Co-suppression" refers to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar foreign or endogenous genes (U.S. Patent No. 5,231,020, incorporated herein by reference).

"Altered levels" refers to the production of gene product(s) in transgenic organisms in amounts or proportions that differ from that of normal or non-transformed organisms.

"Mature" protein refers to a post-translationally processed polypeptide; i.e., one from which any pre- or propeptides present in the primary translation product have been removed.

"Precursor" protein refers to the primary product of translation of mRNA; i.e., with pre- and propeptides still present. Pre- and propeptides may be but are not limited to intracellular localization signals.

A "chloroplast transit peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the chloroplast or other plastid types present in the cell in which the protein is made. "Chloroplast transit sequence" refers to a nucleotide sequence that encodes a chloroplast transit peptide. A "signal peptide" is an amino acid sequence which is translated in conjunction with a protein and directs the protein to the secretory system (Chrispeels, J. J., (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol.* 42:21-53). If the protein is to be directed to a vacuole, a vacuolar targeting signal (*supra*) can further be added, or if to the endoplasmic reticulum, an endoplasmic reticulum retention signal (*supra*) may be added. If the protein is to be directed to the nucleus, any signal peptide present should be removed and instead a nuclear localization signal included (Raikhel (1992) *Plant Phys.* 100:1627-1632).

- "Transformation" refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic" organisms. Examples of methods of plant transformation include Agrobacterium-mediated transformation (De Blaere et al. (1987) *Meth. Enzymol.* 143:277) and particle-accelerated or "gene gun" transformation technology (Klein et al. (1987) *Nature (London)* 327:70-73; U.S. Patent No. 4,945,050).

- Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described more fully in Sambrook, J., Fritsch, E.F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1989 (hereinafter "Maniatis").

- Nucleic acid fragments encoding at least a portion of several phytic acid biosynthetic enzymes have been isolated and identified by comparison of random plant cDNA sequences to public databases containing nucleotide and protein sequences using the BLAST algorithms well known to those skilled in the art. Table 1 lists the proteins that are described herein, and the designation of the cDNA clones that comprise the nucleic acid fragments encoding these proteins.

TABLE 1
Phytic Acid Biosynthetic Enzymes

Enzyme	Clone	Plant
Inositol 1,3,4-trisphosphate 5/6-kinase	cca.pk0022.e6	Corn
	cpe1c.pk001.h8	Corn
	cr1n.pk0188.g8	Corn
	dms2c.pk003.m14	African daisy
	ncs.pk0019.a6	Catalpa
	p0005.cbmf77r	Corn
	p0041.crtcl17r	Corn
	p0090.cspsg46r	Corn
	p0125.czaaj15r	Corn
	p0125.czabp28r	Corn
	p0125.czabp82r	Corn
	rl0n.pk0015.b2	Rice
	rlr12.pk0012.c11	Rice
	rr1.pk0052.f1	Rice
	rr1.pk0061.c5	Rice
	sfl1.pk0091.c9	Soybean
	sgs3n.pk001.b5	Soybean
	sl1.pk0026.a8	Soybean
	sls2c.pk013.j24	Soybean

Enzyme	Clone	Plant
	wdk4c.pk005.a15(5')	Wheat
	wdk4c.pk005.a15(3')	Wheat
	wr1.pk0137.c5(5')	Wheat
	wr1.pk0137.c5(3')	Wheat
	wr1.pk0150.e10	Wheat

The nucleic acid fragments of the instant invention may be used to isolate cDNAs and genes encoding homologous proteins from the same or other plant species. Isolation of homologous genes using sequence-dependent protocols is well known in the art. Examples of sequence-dependent protocols include, but are not limited to, methods of nucleic acid hybridization, and methods of DNA and RNA amplification as exemplified by various uses of nucleic acid amplification technologies (e.g., polymerase chain reaction, ligase chain reaction).

For example, genes encoding other inositol 1,3,4-triphosphate 5/6-kinases, either as cDNAs or genomic DNAs, could be isolated directly by using all or a portion of the instant nucleic acid fragments as DNA hybridization probes to screen libraries from any desired plant employing methodology well known to those skilled in the art. Specific oligonucleotide probes based upon the instant nucleic acid sequences can be designed and synthesized by methods known in the art (Maniatis). Moreover, the entire sequences can be used directly to synthesize DNA probes by methods known to the skilled artisan such as random primer DNA labeling, nick translation, or end-labeling techniques, or RNA probes using available *in vitro* transcription systems. In addition, specific primers can be designed and used to amplify a part or all of the instant sequences. The resulting amplification products can be labeled directly during amplification reactions or labeled after amplification reactions, and used as probes to isolate full length cDNA or genomic fragments under conditions of appropriate stringency.

In addition, two short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols to amplify longer nucleic acid fragments encoding homologous genes from DNA or RNA. The polymerase chain reaction may also be performed on a library of cloned nucleic acid fragments wherein the sequence of one primer is derived from the instant nucleic acid fragments, and the sequence of the other primer takes advantage of the presence of the polyadenylic acid tracts to the 3' end of the mRNA precursor encoding plant genes. Alternatively, the second primer sequence may be based upon sequences derived from the cloning vector. For example, the skilled artisan can follow the RACE protocol (Frohman et al., (1988) *PNAS USA* 85:8998) to generate cDNAs by using PCR to amplify copies of the region between a single point in the transcript and the 3' or 5' end. Primers oriented in the 3' and 5' directions can be designed from the instant

sequences. Using commercially available 3' RACE or 5' RACE systems (BRL), specific 3' or 5' cDNA fragments can be isolated (Ohara et al., (1989) *PNAS USA* 86:5673; Loh et al., (1989) *Science* 243:217). Products generated by the 3' and 5' RACE procedures can be combined to generate full-length cDNAs (Frohman, M. A. and Martin, G. R., (1989)

5 *Techniques* 1:165).

Availability of the instant nucleotide and deduced amino acid sequences facilitates immunological screening of cDNA expression libraries. Synthetic peptides representing portions of the instant amino acid sequences may be synthesized. These peptides can be used to immunize animals to produce polyclonal or monoclonal antibodies with specificity for peptides or proteins comprising the amino acid sequences. These antibodies can be then be used to screen cDNA expression libraries to isolate full-length cDNA clones of interest (Lerner, R. A. (1984) *Adv. Immunol.* 36:1; Maniatis).

The nucleic acid fragments of the instant invention may be used to create transgenic plants in which the disclosed inositol 1,3,4-triphosphate 5/6-kinases are present at higher or lower levels than normal or in cell types or developmental stages in which they are not normally found. This would have the effect of altering the level of inositol 1,3,4-triphosphate 5/6-kinase in those cells.

Overexpression of the inositol 1,3,4-triphosphate 5/6-kinase proteins of the instant invention may be accomplished by first constructing a chimeric gene in which the coding region is operably linked to a promoter capable of directing expression of a gene in the desired tissues at the desired stage of development. For reasons of convenience, the chimeric gene may comprise promoter sequences and translation leader sequences derived from the same genes. 3' Non-coding sequences encoding transcription termination signals may also be provided. The instant chimeric gene may also comprise one or more introns in order to facilitate gene expression.

Plasmid vectors comprising the instant chimeric gene can then be constructed. The choice of plasmid vector is dependent upon the method that will be used to transform host plants. The skilled artisan is well aware of the genetic elements that must be present on the plasmid vector in order to successfully transform, select and propagate host cells containing the chimeric gene. The skilled artisan will also recognize that different independent transformation events will result in different levels and patterns of expression (Jones et al., (1985) *EMBO J.* 4:2411-2418; De Almeida et al., (1989) *Mol. Gen. Genetics* 218:78-86), and thus that multiple events must be screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of DNA, Northern analysis of mRNA expression, Western analysis of protein expression, or phenotypic analysis.

For some applications it may be useful to direct the instant phytyc acid biosynthetic enzymes to different cellular compartments, or to facilitate its secretion from the cell. It is thus envisioned that the chimeric gene described above may be further supplemented by

altering the coding sequence to encode an inositol 1,3,4-triphosphate 5/6-kinase with appropriate intracellular targeting sequences such as transit sequences (Keegstra, K. (1989) *Cell* 56:247-253), signal sequences or sequences encoding endoplasmic reticulum localization (Chrispeels, J.J., (1991) *Ann. Rev. Plant Phys. Plant Mol. Biol.* 42:21-53), or nuclear localization signals (Raikhel, N. (1992) *Plant Phys.* 100:1627-1632) added and/or with targeting sequences that are already present removed. While the references cited give examples of each of these, the list is not exhaustive and more targeting signals of utility may be discovered in the future.

It may also be desirable to reduce or eliminate expression of genes encoding inositol 1,3,4-triphosphate 5/6-kinase in plants for some applications. In order to accomplish this, a chimeric gene designed for co-suppression of the instant phytic acid biosynthetic enzymes can be constructed by linking a gene or gene fragment encoding an inositol 1,3,4-triphosphate 5/6-kinase to plant promoter sequences. Alternatively, a chimeric gene designed to express antisense RNA for all or part of the instant nucleic acid fragment can be constructed by linking the gene or gene fragment in reverse orientation to plant promoter sequences. Either the co-suppression or antisense chimeric genes could be introduced into plants via transformation wherein expression of the corresponding endogenous genes are reduced or eliminated.

The instant inositol 1,3,4-triphosphate 5/6-kinases (or portions thereof) may be produced in heterologous host cells, particularly in the cells of microbial hosts, and can be used to prepare antibodies to these proteins by methods well known to those skilled in the art. The antibodies are useful for detecting inositol 1,3,4-triphosphate 5/6-kinase *in situ* in cells or *in vitro* in cell extracts. Preferred heterologous host cells for production of the instant inositol 1,3,4-triphosphate 5/6-kinases are microbial hosts. Microbial expression systems and expression vectors containing regulatory sequences that direct high level expression of foreign proteins are well known to those skilled in the art. Any of these could be used to construct a chimeric gene for production of the instant inositol 1,3,4-triphosphate 5/6-kinases. This chimeric gene could then be introduced into appropriate microorganisms via transformation to provide high level expression of the encoded phytic acid biosynthetic enzyme. An example of a vector for high level expression of the instant inositol 1,3,4-triphosphate 5/6-kinases in a bacterial host is provided (Example 6).

All or a substantial portion of the nucleic acid fragments of the instant invention may also be used as probes for genetically and physically mapping the genes that they are a part of, and as markers for traits linked to those genes. Such information may be useful in plant breeding in order to develop lines with desired phenotypes. For example, the instant nucleic acid fragments may be used as restriction fragment length polymorphism (RFLP) markers. Southern blots (Maniatis) of restriction-digested plant genomic DNA may be probed with the nucleic acid fragments of the instant invention. The resulting banding patterns may then be subjected to genetic analyses using computer programs such as MapMaker (Lander et al.,

(1987) *Genomics* 1:174-181) in order to construct a genetic map. In addition, the nucleic acid fragments of the instant invention may be used to probe Southern blots containing restriction endonuclease-treated genomic DNAs of a set of individuals representing parent and progeny of a defined genetic cross. Segregation of the DNA polymorphisms is noted and used to calculate the position of the instant nucleic acid sequence in the genetic map previously obtained using this population (Botstein, D. et al., (1980) *Am. J. Hum. Genet.* 32:314-331).

The production and use of plant gene-derived probes for use in genetic mapping is described in R. Bernatzky, R. and Tanksley, S. D. (1986) *Plant Mol. Biol. Reporter* 4(1):37-41. Numerous publications describe genetic mapping of specific cDNA clones using the methodology outlined above or variations thereof. For example, F2 intercross populations, backcross populations, randomly mated populations, near isogenic lines, and other sets of individuals may be used for mapping. Such methodologies are well known to those skilled in the art.

Nucleic acid probes derived from the instant nucleic acid sequences may also be used for physical mapping (i.e., placement of sequences on physical maps; see Hoheisel, J. D., et al., In: *Nonmammalian Genomic Analysis: A Practical Guide*, Academic press 1996, pp. 319-346, and references cited therein).

In another embodiment, nucleic acid probes derived from the instant nucleic acid sequences may be used in direct fluorescence *in situ* hybridization (FISH) mapping (Trask, B. J. (1991) *Trends Genet.* 7:149-154). Although current methods of FISH mapping favor use of large clones (several to several hundred KB; see Laan, M. et al. (1995) *Genome Research* 5:13-20), improvements in sensitivity may allow performance of FISH mapping using shorter probes.

A variety of nucleic acid amplification-based methods of genetic and physical mapping may be carried out using the instant nucleic acid sequences. Examples include allele-specific amplification (Kazazian, H. H. (1989) *J. Lab. Clin. Med.* 114(2):95-96), polymorphism of PCR-amplified fragments (CAPS; Sheffield, V. C. et al. (1993) *Genomics* 16:325-332), allele-specific ligation (Landegren, U. et al. (1988) *Science* 241:1077-1080), nucleotide extension reactions (Sokolov, B. P. (1990) *Nucleic Acid Res.* 18:3671), Radiation Hybrid Mapping (Walter, M. A. et al. (1997) *Nature Genetics* 7:22-28) and Happy Mapping (Dear, P. H. and Cook, P. R. (1989) *Nucleic Acid Res.* 17:6795-6807). For these methods, the sequence of a nucleic acid fragment is used to design and produce primer pairs for use in the amplification reaction or in primer extension reactions. The design of such primers is well known to those skilled in the art. In methods employing PCR-based genetic mapping, it may be necessary to identify DNA sequence differences between the parents of the mapping cross in the region corresponding to the instant nucleic acid sequence. This, however, is generally not necessary for mapping methods.

- Loss of function mutant phenotypes may be identified for the instant cDNA clones either by targeted gene disruption protocols or by identifying specific mutants for these genes contained in a maize population carrying mutations in all possible genes (Ballinger and Benzer, (1989) *Proc. Natl. Acad. Sci USA* 86:9402; Koes et al., (1995) *Proc. Natl. Acad. Sci USA* 92:8149; Bensen et al., (1995) *Plant Cell* 7:75). The latter approach may be accomplished in two ways. First, short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols in conjunction with a mutation tag sequence primer on DNAs prepared from a population of plants in which Mutator transposons or some other mutation-causing DNA element has been introduced (see Bensen, *supra*). The
- 10 amplification of a specific DNA fragment with these primers indicates the insertion of the mutation tag element in or near the plant gene encoding the inositol 1,3,4-triphosphate 5/6-kinase. Alternatively, the instant nucleic acid fragment may be used as a hybridization probe against PCR amplification products generated from the mutation population using the mutation tag sequence primer in conjunction with an arbitrary genomic site primer. such as
- 15 that for a restriction enzyme site-anchored synthetic adaptor. With either method, a plant containing a mutation in the endogenous gene encoding an inositol 1,3,4-triphosphate 5/6-kinase can be identified and obtained. This mutant plant can then be used to determine or confirm the natural function of the inositol 1,3,4-triphosphate 5/6-kinase gene product.

EXAMPLES

- 20 The present invention is further defined in the following Examples, in which all parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without
- 25 departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

EXAMPLE 1

Composition of cDNA Libraries: Isolation and Sequencing of cDNA Clones

- 30 cDNA libraries representing mRNAs from various african daisy, *Catalpa*, corn, rice, soybean and wheat tissues were prepared. The characteristics of the libraries are described below.

TABLE 2

cDNA Libraries from African Daisy, *Catalpa*, Corn, Rice, Soybean and Wheat

Library	Tissue	Clone
cca	Corn (<i>Zea mays</i> L.) type II callus tissue, undifferentiated, highly transformable	cca.pk0022.e6
cpe1c	Corn (<i>Zea mays</i> L.) pooled BMS treated with chemicals related to phosphatase***	cpe1c.pk001.h8
cr1n	Corn (<i>Zea mays</i> L.) root from 7 day seedlings grown in light*	cr1n.pk0188.g8
dms2c	African daisy (<i>Dimorphotheca sinuata</i>) developing seeds	dms2c.pk003.m14
ncs	<i>Catalpa speciosa</i> developing seed	ncs.pk0019.a6
p0005	Corn (<i>Zea mays</i> L.) immature ear	p0005.cbmf77r
p0041	Corn (<i>Zea mays</i> L.) root tips (four days after imbibition), smaller than 5 mm in length.	p0041.crtcl17r
p0090	Corn (<i>Zea mays</i> L.) heat shocked seedling after 10 day drought stress (heat shocked for 8, 16, 24 hours at 45C) pooled for library construction*	p0090.cspsg46r
p0125	Corn (<i>Zea mays</i> L.) anther: prophase I	p0125.czaaj15r p0125.czabg28r p0125.czabp82r
rl0n	Rice (<i>Oryza sativa</i> L.) 15 day leaf*	rl0n.pk0015.b2
rlr12	Rice (<i>Oryza sativa</i> L.) leaf, 15 days after germination, 12 hours after infection of <i>Magaporthe grisea</i> strain 4360-R-62 (AVR2-YAMO)	rlr12.pk0012.c11
rr1	Rice (<i>Oryza sativa</i> L.) root of two week old developing seedling	rr1.pk0052.f1 rr1.pk0061.c5
sfl1	Soybean (<i>Glycine max</i> L.) immature flower	sfl1.pk0091.c9
sgs3n	Soybean (<i>Glycine max</i> L.) seeds 25 hrs after germination	sgs3n.pk001.b5
sl1	Soybean (<i>Glycine max</i> L.) two week old developing seedlings treated with water	sl1.pk0026.a8
sls2c	Soybean (<i>Glycine max</i> L.) infected with <i>Sclerotinia sclerotiorum</i> mycelium	sls2c.pk013.j24
wdk4c	Wheat (<i>Triticum aestivum</i> L.) developing kernel, 21 days after anthesis	wdk4c.pk005.a15(5') wdk4c.pk005.a15(3')
wr1	Wheat (<i>Triticum aestivum</i> L.) root; 7 day old seedling, light grown	wr1.pk0137.c5(5') wr1.pk0137.c5(3') wr1.pk0150.e10

*These libraries were normalized essentially as described in U.S. Patent No. 5,482,845

**V-12 refers to stages of corn growth. The descriptions can be found in "How a Corn

Plant Develops" Special Report No.48, Iowa State University of Science and Technology Cooperative Extension Service Ames,Iowa, Reprinted February 1996.

***Chemicals related to phosphatase: okadaic acid, cyclosporin A, calyculin A and cypermethrin Sigma Chemical Co. and Calbiochem-Novabiochem Corp Calbiochem.

cDNA libraries were prepared in Uni-ZAP™ XR vectors according to the manufacturer's protocol (Stratagene Cloning Systems, La Jolla, CA). Conversion of the Uni-ZAP™ XR libraries into plasmid libraries was accomplished according to the protocol provided by Stratagene. Upon conversion, cDNA inserts were contained in the plasmid vector pBluescript. cDNA inserts from randomly picked bacterial colonies containing recombinant pBluescript plasmids were amplified via polymerase chain reaction using primers specific for vector sequences flanking the inserted cDNA sequences or plasmid DNA was prepared from cultured bacterial cells. Amplified insert DNAs or plasmid DNAs were sequenced in dye-primer sequencing reactions to generate partial cDNA sequences (expressed sequence tags or "ESTs"; see Adams, M. D. et al., (1991) *Science* 252:1651). The resulting ESTs were analyzed using a Perkin Elmer Model 377 fluorescent sequencer.

EXAMPLE 2

Identification of cDNA Clones

ESTs encoding phytic acid biosynthetic enzymes were identified by conducting BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al., (1993) *J. Mol. Biol.* 215:403-410; see also www.ncbi.nlm.nih.gov/BLAST/) searches for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the last major release of the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The cDNA sequences obtained in Example 1 were analyzed for similarity to all publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm provided by the National Center for Biotechnology Information (NCBI). The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish, W. and States, D. J. (1993) *Nature Genetics* 3:266-272 and Altschul, Stephen F., et al. (1997) *Nucleic Acids Res.* 25:3389-3402) provided by the NCBI. For convenience, the P-value (probability) of observing a match of a cDNA sequence to a sequence contained in the searched databases merely by chance as calculated by BLAST are reported herein as "pLog" values, which represent the negative of the logarithm of the reported P-value. Accordingly, the greater the pLog value, the greater the likelihood that the cDNA sequence and the BLAST "hit" represent homologous proteins.

EXAMPLE 3

Characterization of cDNA Clones Encoding Inositol 1,3,4-Triphosphate 5/6-Kinase

The BLASTX search using the EST sequences from clones cca.pk0022.e6, cpe1.c.pk001.h8, cr1n.pk0188.g8, dms2c.pk003.m14, rlr12.pk0012.c11, rr1.pk0061.c5.

sf11.pk0091.c9, sls2c.pk013.j24, wdk4c.pk005.a15(5'), wdk4c.pk005.a15(3'), wr1.pk0137.c5(5'), wr1.pk0137.c5(3') and wr1.pk0150.e10 revealed similarity of the proteins encoded by the cDNAs to inositol 1,3,4-triphosphate 5/6-kinase from *Arabidopsis thaliana* (NCBI Identifier No. gi 3396079).

- 5 The BLASTX search using the EST sequences from clones ncs.pk0019.a6, p0125.czaaj15r, p0125.czabg28r, p0125.czabp82r, p0041.crtcl17r, rr1.pk0052.fl, rl0n.pk0015.b2, sgs3n.pk001.b5 and sl1.pk0026.a8 revealed similarity of the proteins encoded by the cDNAs to inositol 1,3,4-triphosphate 5/6-kinase from *Arabidopsis thaliana* (NCBI Identifier No. gi 3660465).
- 10 In the process of comparing the ESTs it was found that several had overlapping regions of homology. Using this homology it was possible to align the ESTs and assemble several contigs encoding unique inositol 1,3,4-triphosphate 5/6-kinase proteins. The composition of each of the assembled contigs is shown in Table 3.

The BLAST results for each of the ESTs and the contigs are also shown in Table 3:

15

TABLE 3

BLAST Results for Clones Encoding Polypeptides Homologous
to *Arabidopsis thaliana* Inositol 1,3,4-Triphosphate 5/6-Kinase

Clone	BLAST pLog Score
Contig composed of clones:	67.40
cca.pk0022.e6	
cpe1c.pk001.h8	
cr1n.pk0188.g8	
p0005.cbmtfp77r	
p0090.cspsg46r	
dms2c.pk003.m14	25.15
ncs.pk0019.a6	48.52
Contig composed of clones:	74.05
p0125.czaaj15r	
p0125.czabg28r	
p0125.czabp82r	
p0041.crtcl17r	
Contig composed of	44.15
rr1.pk0052.fl	
rl0n.pk0015.b2	
rlr12.pk0012.c11	96.22
rr1.pk0061.c5	21.30
sf11.pk0091.c9	107.00
sgs3n.pk001.b5	72.52
sl1.pk0026.a8	24.70
sls2c.pk013.j24	102.00

Clone	BLAST pLog Score
wdk4c.pk005.a15(3')	21.10
wdk4c.pk005.a15(3')	15.52
wrl.pk0137.c5(5')	29.00
wrl.pk0137.c5(3')	6.70
wrl.pk0150.e10	59.00

The sequence of the corn contig composed of clones cca.pk0022.e6, cpe1c.pk001.h8, cr1n.pk0188.g8, p0005.cbmf77r and p0090.cpsg46r is shown in SEQ ID NO:1; the deduced amino acid sequence of this cDNA, which represents 100% of the of the protein, is shown in SEQ ID NO:2. The amino acid sequence set forth in SEQ ID NO:2 was evaluated by BLASTP, yielding a pLog value of 59.22 versus the *Arabidopsis thaliana* sequence. A calculation of the percent similarity of the amino acid sequence set forth in SEQ ID NO:2 and the *Arabidopsis thaliana* sequence (using the Clustal algorithm) revealed that the protein encoded by SEQ ID NO:2 is 37% similar to the *Arabidopsis thaliana* inositol 1,3,4-triphosphate 5/6-kinase.

The sequence of a portion of the cDNA insert from clone dms2c.pk003.m14 is shown in SEQ ID NO:3; the deduced amino acid sequence of this cDNA, which represents 35% of the of the protein (N-terminal region), is shown in SEQ ID NO:4. A calculation of the percent similarity of the amino acid sequence set forth in SEQ ID NO:4 and the *Arabidopsis thaliana* sequence (using the Clustal algorithm) revealed that the protein encoded by SEQ ID NO:4 is 37% similar to the *Arabidopsis thaliana* inositol 1,3,4-triphosphate 5/6-kinase.

The sequence of a portion of the cDNA insert from clone ncs.pk0019.a6 is shown in SEQ ID NO:5; the deduced amino acid sequence of this cDNA, which represents 73% of the of the protein (C-terminal region), is shown in SEQ ID NO:6. The amino acid sequence set forth in SEQ ID NO:6 was evaluated by BLASTP, yielding a pLog value of 46.52 versus the *Arabidopsis thaliana* sequence. A calculation of the percent similarity of the amino acid sequence set forth in SEQ ID NO:6 and the *Arabidopsis thaliana* sequence (using the Clustal algorithm) revealed that the protein encoded by SEQ ID NO:6 is 39% similar to the *Arabidopsis thaliana* inositol 1,3,4-triphosphate 5/6-kinase.

The sequence of the corn contig composed of clones p0125.czaaj15r, p0125.czabg28r, p0125.czabp82r and p0041.crtcl17r is shown in SEQ ID NO:7; the deduced amino acid sequence of this cDNA, which represents 100% of the of the protein, is shown in SEQ ID NO:8. The amino acid sequence set forth in SEQ ID NO:8 was evaluated by BLASTP, yielding a pLog value of 84.40 versus the *Arabidopsis thaliana* sequence. A calculation of the percent similarity of the amino acid sequence set forth in SEQ ID NO:8 and the *Arabidopsis thaliana* sequence (using the Clustal algorithm) revealed that the protein

encoded by SEQ ID NO:8 is 48% similar to the *Arabidopsis thaliana* inositol 1,3,4-triphosphate 5/6-kinase.

The sequence of the rice contig composed of clones rr1.pk0052.f1 and ri0n.pk0015.b2 is shown in SEQ ID NO:9; the deduced amino acid sequence of this cDNA, which represents 77% of the of the protein (C-terminal region), is shown in SEQ ID NO:10. The amino acid sequence set forth in SEQ ID NO:10 was evaluated by BLASTP, yielding a pLog value of 43.70 versus the *Arabidopsis thaliana* sequence. A calculation of the percent similarity of the amino acid sequence set forth in SEQ ID NO:10 and the *Arabidopsis thaliana* sequence (using the Clustal algorithm) revealed that the protein encoded by SEQ ID NO:10 is 37% similar to the *Arabidopsis thaliana* inositol 1,3,4-triphosphate 5/6-kinase.

The sequence of the entire cDNA insert from clone rlr12.pk0012.c11 is shown in SEQ ID NO:11; the deduced amino acid sequence of this cDNA, which represents 100% of the of the protein, is shown in SEQ ID NO:12. The amino acid sequence set forth in SEQ ID NO:12 was evaluated by BLASTP, yielding a pLog value of 96.40 versus the *Arabidopsis thaliana* sequence. A calculation of the percent similarity of the amino acid sequence set forth in SEQ ID NO:12 and the *Arabidopsis thaliana* sequence (using the Clustal algorithm) revealed that the protein encoded by SEQ ID NO:12 is 53% similar to the *Arabidopsis thaliana* inositol 1,3,4-triphosphate 5/6-kinase.

The sequence of a portion of the cDNA insert from clone rr1.pk0061.c5 is shown in SEQ ID NO:13; the deduced amino acid sequence of this cDNA, which represents 41% of the of the protein (N-terminal region), is shown in SEQ ID NO:14. A calculation of the percent similarity of the amino acid sequence set forth in SEQ ID NO:14 and the *Arabidopsis thaliana* sequence (using the Clustal algorithm) revealed that the protein encoded by SEQ ID NO:14 is 36% similar to the *Arabidopsis thaliana* inositol 1,3,4-triphosphate 5/6-kinase.

The sequence of the entire cDNA insert from clone sfl1.pk0091.c9 is shown in SEQ ID NO:15; the deduced amino acid sequence of this cDNA, which represents 100% of the of the protein, is shown in SEQ ID NO:16. The amino acid sequence set forth in SEQ ID NO:16 was evaluated by BLASTP, yielding a pLog value of 94.70 versus the *Arabidopsis thaliana* sequence. A calculation of the percent similarity of the amino acid sequence set forth in SEQ ID NO:16 and the *Arabidopsis thaliana* sequence (using the Clustal algorithm) revealed that the protein encoded by SEQ ID NO:16 is 53% similar to the *Arabidopsis thaliana* inositol 1,3,4-triphosphate 5/6-kinase.

The sequence of the entire cDNA insert from clone sgs3n.pk001.b5 is shown in SEQ ID NO:17; the deduced amino acid sequence of this cDNA, which represents 100% of the of the protein, is shown in SEQ ID NO:18. The amino acid sequence set forth in SEQ ID NO:18 was evaluated by BLASTP, yielding a pLog value of 67.00 versus the *Arabidopsis thaliana* sequence. A calculation of the percent similarity of the amino acid sequence set forth in SEQ ID NO:18 and the *Arabidopsis thaliana* sequence (using the Clustal algorithm)

revealed that the protein encoded by SEQ ID NO:18 is 40% similar to the *Arabidopsis thaliana* inositol 1,3,4-triphosphate 5/6-kinase.

The sequence of a portion of the cDNA insert from clone sl1.pk0026.a8 is shown in SEQ ID NO:19; the deduced amino acid sequence of this cDNA, which represents 41% of the of the protein (C-terminal region), is shown in SEQ ID NO:20. The amino acid sequence set forth in SEQ ID NO:20 was evaluated by BLASTP, yielding a pLog value of 22.70 versus the *Arabidopsis thaliana* sequence. A calculation of the percent similarity of the amino acid sequence set forth in SEQ ID NO:20 and the *Arabidopsis thaliana* sequence (using the Clustal algorithm) revealed that the protein encoded by SEQ ID NO:20 is 40% similar to the *Arabidopsis thaliana* inositol 1,3,4-triphosphate 5/6-kinase.

The sequence of the entire cDNA insert from clone sls2c.pk013.j24 is shown in SEQ ID NO:21; the deduced amino acid sequence of this cDNA, which represents 99% of the of the protein, is shown in SEQ ID NO:22. The amino acid sequence set forth in SEQ ID NO:22 was evaluated by BLASTP, yielding a pLog value of 92.00 versus the *Arabidopsis thaliana* sequence. A calculation of the percent similarity of the amino acid sequence set forth in SEQ ID NO:22 and the *Arabidopsis thaliana* sequence (using the Clustal algorithm) revealed that the protein encoded by SEQ ID NO:22 is 51% similar to the *Arabidopsis thaliana* inositol 1,3,4-triphosphate 5/6-kinase.

The sequence of a portion of the cDNA insert from clone wdk4c.pk005.a15(5') is shown in SEQ ID NO:23; the deduced amino acid sequence of this cDNA, which represents 51% of the of the protein (N-terminal region), is shown in SEQ ID NO:24. A calculation of the percent similarity of the amino acid sequence set forth in SEQ ID NO:24 and the *Arabidopsis thaliana* sequence (using the Clustal algorithm) revealed that the protein encoded by SEQ ID NO:24 is 32% similar to the *Arabidopsis thaliana* inositol 1,3,4-triphosphate 5/6-kinase.

The sequence of a portion of the cDNA insert from clone wdk4c.pk005.a15(3') is shown in SEQ ID NO:25; the deduced amino acid sequence of this cDNA, which represents 31% of the of the protein (C-terminal region), is shown in SEQ ID NO:26. A calculation of the percent similarity of the amino acid sequence set forth in SEQ ID NO:26 and the *Arabidopsis thaliana* sequence (using the Clustal algorithm) revealed that the protein encoded by SEQ ID NO:26 is 29% similar to the *Arabidopsis thaliana* inositol 1,3,4-triphosphate 5/6-kinase.

The sequence of a portion of the cDNA insert from clone wr1.pk0137.c5(5') is shown in SEQ ID NO:27; the deduced amino acid sequence of this cDNA, which represents 50% of the of the protein (middle region), is shown in SEQ ID NO:28. A calculation of the percent similarity of the amino acid sequence set forth in SEQ ID NO:28 and the *Arabidopsis thaliana* sequence (using the Clustal algorithm) revealed that the protein encoded by SEQ ID NO:28 is 40% similar to the *Arabidopsis thaliana* inositol 1,3,4-triphosphate 5/6-kinase.

The sequence of a portion of the cDNA insert from clone wr1.pk0137.c5(3') is shown in SEQ ID NO:29; the deduced amino acid sequence of this cDNA, which represents 21% of the of the protein (C-terminal region), is shown in SEQ ID NO:30. A calculation of the percent similarity of the amino acid sequence set forth in SEQ ID NO:30 and the *Arabidopsis thaliana* sequence (using the Clustal algorithm) revealed that the protein encoded by SEQ ID NO:30 is 35% similar to the *Arabidopsis thaliana* inositol 1,3,4-triphosphate 5/6-kinase.

The sequence of a portion of the cDNA insert from clone wr1.pk0150.e10 is shown in SEQ ID NO:31; the deduced amino acid sequence of this cDNA, which represents 88% of the of the protein (C-terminal region), is shown in SEQ ID NO:32. The amino acid sequence set forth in SEQ ID NO:32 was evaluated by BLASTP, yielding a plog value of 51.15 versus the *Arabidopsis thaliana* sequence. A calculation of the percent similarity of the amino acid sequence set forth in SEQ ID NO:32 and the *Arabidopsis thaliana* sequence (using the Clustal algorithm) revealed that the protein encoded by SEQ ID NO:32 is 38% similar to the *Arabidopsis thaliana* inositol 1,3,4-triphosphate 5/6-kinase.

Figure 1 presents an alignment of the amino acid sequence set forth in SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30 and 32 with the *Arabidopsis thaliana* inositol 1,3,4-triphosphate 5/6-kinase amino acid sequences, SEQ ID NO:33 (NCBI Identifier No. gi 3396079) and SEQ ID NO:34 (NCBI Identifier No. gi 3660465).

Alignments were performed using the Clustal algorithm.

These sequences represent the first african daisy, *Catalpa*, corn, rice, soybean and wheat sequences encoding inositol 1,3,4-triphosphate 5/6-kinase proteins.

EXAMPLE 4

Expression of Chimeric Genes in Monocot Cells

A chimeric gene comprising a cDNA encoding phytyc acid biosynthetic enzyme in sense orientation with respect to the maize 27 kD zein promoter that is located 5' to the cDNA fragment, and the 10 kD zein 3' end that is located 3' to the cDNA fragment, can be constructed. The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites (NcoI or SmaI) can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the digested vector pML103 as described below. Amplification is then performed in a standard PCR. The amplified DNA is then digested with restriction enzymes NcoI and SmaI and fractionated on an agarose gel. The appropriate band can be isolated from the gel and combined with a 4.9 kb NcoI-SmaI fragment of the plasmid pML103. Plasmid pML103 has been deposited under the terms of the Budapest Treaty at ATCC (American Type Culture Collection, 10801 University Blvd., Manassas, VA 20110-2209), and bears accession number ATCC 97366. The DNA segment from pML103 contains a 1.05 kb SalI-NcoI promoter fragment of the maize 27 kD zein gene and a 0.96 kb SmaI-SalI fragment from the 3' end of the maize 10 kD zein gene in the vector pGem9Zf(+)

(Promega). Vector and insert DNA can be ligated at 15°C overnight, essentially as described (Maniatis). The ligated DNA may then be used to transform *E. coli* XL1-Blue (Epicurian Coli XL-1 Blue™; Stratagene). Bacterial transformants can be screened by restriction enzyme digestion of plasmid DNA and limited nucleotide sequence analysis using the dideoxy chain termination method (Sequenase™ DNA Sequencing Kit; U.S. Biochemical). The resulting plasmid construct would comprise a chimeric gene encoding, in the 5' to 3' direction, the maize 27 kD zein promoter, a cDNA fragment encoding a phytic acid biosynthetic enzymes, and the 10 kD zein 3' region.

The chimeric gene described above can then be introduced into corn cells by the following procedure. Immature corn embryos can be dissected from developing caryopses derived from crosses of the inbred corn lines H99 and LH132. The embryos are isolated 10 to 11 days after pollination when they are 1.0 to 1.5 mm long. The embryos are then placed with the axis-side facing down and in contact with agarose-solidified N6 medium (Chu et al., (1975) *Sci. Sin. Peking* 18:659-668). The embryos are kept in the dark at 27°C. Friable embryogenic callus consisting of undifferentiated masses of cells with somatic proembryoids and embryoids borne on suspensor structures proliferates from the scutellum of these immature embryos. The embryogenic callus isolated from the primary explant can be cultured on N6 medium and sub-cultured on this medium every 2 to 3 weeks.

The plasmid, p35S/Ac (obtained from Dr. Peter Eckes, Hoechst Ag, Frankfurt, Germany) may be used in transformation experiments in order to provide for a selectable marker. This plasmid contains the *Pat* gene (see European Patent Publication 0 242 236) which encodes phosphinothricin acetyl transferase (PAT). The enzyme PAT confers resistance to herbicidal glutamine synthetase inhibitors such as phosphinothricin. The *pat* gene in p35S/Ac is under the control of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*.

The particle bombardment method (Klein et al., (1987) *Nature* 327:70-73) may be used to transfer genes to the callus culture cells. According to this method, gold particles (1 µm in diameter) are coated with DNA using the following technique. Ten µg of plasmid DNAs are added to 50 µL of a suspension of gold particles (60 mg per mL). Calcium chloride (50 µL of a 2.5 M solution) and spermidine free base (20 µL of a 1.0 M solution) are added to the particles. The suspension is vortexed during the addition of these solutions. After 10 minutes, the tubes are briefly centrifuged (5 sec at 15,000 rpm) and the supernatant removed. The particles are resuspended in 200 µL of absolute ethanol, centrifuged again and the supernatant removed. The ethanol rinse is performed again and the particles resuspended in a final volume of 30 µL of ethanol. An aliquot (5 µL) of the DNA-coated gold particles can be placed in the center of a Kapton™ flying disc (Bio-Rad Labs). The particles are then accelerated into the corn tissue with a Biolistic™ PDS-1000/He (Bio-Rad

Instruments, Hercules CA), using a helium pressure of 1000 psi, a gap distance of 0.5 cm and a flying distance of 1.0 cm.

For bombardment, the embryogenic tissue is placed on filter paper over agarose-solidified N6 medium. The tissue is arranged as a thin lawn and covered a circular area of about 5 cm in diameter. The petri dish containing the tissue can be placed in the chamber of the PDS-1000/He approximately 8 cm from the stopping screen. The air in the chamber is then evacuated to a vacuum of 28 inches of Hg. The macrocarrier is accelerated with a helium shock wave using a rupture membrane that bursts when the He pressure in the shock tube reaches 1000 psi.

Seven days after bombardment the tissue can be transferred to N6 medium that contains glufosinate (2 mg per liter) and lacks casein or proline. The tissue continues to grow slowly on this medium. After an additional 2 weeks the tissue can be transferred to fresh N6 medium containing glufosinate. After 6 weeks, areas of about 1 cm in diameter of actively growing callus can be identified on some of the plates containing the glufosinate-supplemented medium. These calli may continue to grow when sub-cultured on the selective medium.

Plants can be regenerated from the transgenic callus by first transferring clusters of tissue to N6 medium supplemented with 0.2 mg per liter of 2,4-D. After two weeks the tissue can be transferred to regeneration medium (Fromm et al., (1990) *Bio/Technology* 8:833-839).

EXAMPLE 5

Expression of Chimeric Genes in Dicot Cells

A seed-specific expression cassette composed of the promoter and transcription terminator from the gene encoding the β subunit of the seed storage protein phaseolin from the bean *Phaseolus vulgaris* (Doyle et al. (1986) *J. Biol. Chem.* 261:9228-9238) can be used for expression of the instant phytic acid biosynthetic enzymes in transformed soybean. The phaseolin cassette includes about 500 nucleotides upstream (5') from the translation initiation codon and about 1650 nucleotides downstream (3') from the translation stop codon of phaseolin. Between the 5' and 3' regions are the unique restriction endonuclease sites Nco I (which includes the ATG translation initiation codon), Sma I, Kpn I and Xba I. The entire cassette is flanked by Hind III sites.

The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the expression vector. Amplification is then performed as described above, and the isolated fragment is inserted into a pUC18 vector carrying the seed expression cassette.

Soybean embryos may then be transformed with the expression vector comprising a sequence encoding a phytic acid biosynthetic enzyme. To induce somatic embryos,

cotyledons, 3-5 mm in length dissected from surface sterilized, immature seeds of the soybean cultivar A2872, can be cultured in the light or dark at 26°C on an appropriate agar medium for 6-10 weeks. Somatic embryos which produce secondary embryos are then excised and placed into a suitable liquid medium. After repeated selection for clusters of somatic embryos which multiplied as early, globular staged embryos, the suspensions are maintained as described below.

Soybean embryogenic suspension cultures can be maintained in 35 mL liquid media on a rotary shaker, 150 rpm, at 26°C with fluorescent lights on a 16:8 hour day/night schedule. Cultures are subcultured every two weeks by inoculating approximately 35 mg of tissue into 35 mL of liquid medium.

Soybean embryogenic suspension cultures may then be transformed by the method of particle gun bombardment (Kline et al. (1987) *Nature* (London) 327:70, U.S. Patent No. 4,945,050). A DuPont Biolistic™ PDS1000/HE instrument (helium retrofit) can be used for these transformations.

A selectable marker gene which can be used to facilitate soybean transformation is a chimeric gene composed of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812), the hygromycin phosphotransferase gene from plasmid pJR225 (from *E. coli*; Grütz et al. (1983) *Gene* 25:179-188) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*. The seed expression cassette comprising the phaseolin 5' region, the fragment encoding the phytic acid biosynthetic enzyme and the phaseolin 3' region can be isolated as a restriction fragment. This fragment can then be inserted into a unique restriction site of the vector carrying the marker gene.

To 50 µL of a 60 mg/mL 1 µm gold particle suspension is added (in order): 5 µL DNA (1 µg/µL), 20 µL spermidine (0.1 M), and 50 µL CaCl₂ (2.5 M). The particle preparation is then agitated for three minutes, spun in a microfuge for 10 seconds and the supernatant removed. The DNA-coated particles are then washed once in 400 µL 70% ethanol and resuspended in 40 µL of anhydrous ethanol. The DNA/particle suspension can be sonicated three times for one second each. Five µL of the DNA-coated gold particles are then loaded on each macro carrier disk.

Approximately 300-400 mg of a two-week-old suspension culture is placed in an empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette. For each transformation experiment, approximately 5-10 plates of tissue are normally bombarded. Membrane rupture pressure is set at 1100 psi and the chamber is evacuated to a vacuum of 28 inches mercury. The tissue is placed approximately 3.5 inches away from the retaining screen and bombarded three times. Following bombardment, the tissue can be divided in half and placed back into liquid and cultured as described above.

Five to seven days post bombardment, the liquid media may be exchanged with fresh media, and eleven to twelve days post bombardment with fresh media containing 50 mg/mL

- hygromycin. This selective media can be refreshed weekly. Seven to eight weeks post bombardment, green, transformed tissue may be observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue is removed and inoculated into individual flasks to generate new, clonally propagated, transformed embryogenic suspension cultures. Each new line may be treated as an independent transformation event. These suspensions can then be subcultured and maintained as clusters of immature embryos or regenerated into whole plants by maturation and germination of individual somatic embryos.

EXAMPLE 6

Expression of Chimeric Genes in Microbial Cells

- The cDNAs encoding the instant phytic acid biosynthetic enzymes can be inserted into the T7 *E. coli* expression vector pBT430. This vector is a derivative of pET-3a (Rosenberg et al. (1987) *Gene* 56:125-135) which employs the bacteriophage T7 RNA polymerase/T7 promoter system. Plasmid pBT430 was constructed by first destroying the EcoR I and Hind III sites in pET-3a at their original positions. An oligonucleotide adaptor containing EcoR I and Hind III sites was inserted at the BamH I site of pET-3a. This created pET-3aM with additional unique cloning sites for insertion of genes into the expression vector. Then, the Nde I site at the position of translation initiation was converted to an Nco I site using oligonucleotide-directed mutagenesis. The DNA sequence of pET-3aM in this region, 5'-CATATGG, was converted to 5'-CCCATGG in pBT430.
- Plasmid DNA containing a cDNA may be appropriately digested to release a nucleic acid fragment encoding the protein. This fragment may then be purified on a 1% NuSieve GTG™ low melting agarose gel (FMC). Buffer and agarose contain 10 µg/ml ethidium bromide for visualization of the DNA fragment. The fragment can then be purified from the agarose gel by digestion with GELase™ (Epicentre Technologies) according to the manufacturer's instructions, ethanol precipitated, dried and resuspended in 20 µL of water. Appropriate oligonucleotide adapters may be ligated to the fragment using T4 DNA ligase (New England Biolabs, Beverly, MA). The fragment containing the ligated adapters can be purified from the excess adapters using low melting agarose as described above. The vector pBT430 is digested, dephosphorylated with alkaline phosphatase (NEB) and deproteinized with phenol/chloroform as described above. The prepared vector pBT430 and fragment can then be ligated at 16°C for 15 hours followed by transformation into DH5 electrocompetent cells (GIBCO BRL). Transformants can be selected on agar plates containing LB media and 100 µg/mL ampicillin. Transformants containing the gene encoding the phytic acid biosynthetic enzyme are then screened for the correct orientation with respect to the T7 promoter by restriction enzyme analysis.

For high level expression, a plasmid clone with the cDNA insert in the correct orientation relative to the T7 promoter can be transformed into *E. coli* strain BL21(DE3) (Studier et al. (1986) *J. Mol. Biol.* 189:113-130). Cultures are grown in LB medium containing ampicillin (100 mg/L) at 25°C. At an optical density at 600 nm of approximately

- 1, IPTG (isopropylthio- β -galactoside, the inducer) can be added to a final concentration of 0.4 mM and incubation can be continued for 3 h at 25°. Cells are then harvested by centrifugation and re-suspended in 50 μ L of 50 mM Tris-HCl at pH 8.0 containing 0.1 mM DTT and 0.2 mM phenyl methylsulfonyl fluoride. A small amount of 1 mm glass beads can be added and the mixture sonicated 3 times for about 5 seconds each time with a microprobe sonicator. The mixture is centrifuged and the protein concentration of the supernatant determined. One μ g of protein from the soluble fraction of the culture can be separated by SDS-polyacrylamide gel electrophoresis. Gels can be observed for protein bands migrating at the expected molecular weight.

10

CLAIMS

What is claimed is:

1. An isolated nucleic acid fragment encoding all or a substantial portion of an inositol 1,3,4-triphosphate 5/6-kinase comprising a member selected from the group consisting of:
 - (a) an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30 and 32;
 - (b) an isolated nucleic acid fragment that is substantially similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30 and 32; and
 - (c) an isolated nucleic acid fragment that is complementary to (a) or (b).
2. The isolated nucleic acid fragment of Claim 1 wherein the nucleotide sequence of the fragment comprises all or a portion of the sequence set forth in a member selected from the group consisting of SEQ ID NO:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29 and 31.
3. A chimeric gene comprising the nucleic acid fragment of Claim 1 operably linked to suitable regulatory sequences.
4. A transformed host cell comprising the chimeric gene of Claim 3.
5. An inositol 1,3,4-triphosphate 5/6-kinase polypeptide comprising all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30 and 32.
6. A method of altering the level of expression of a phytic acid biosynthetic enzyme in a host cell comprising:
 - (a) transforming a host cell with the chimeric gene of Claim 3; and
 - (b) growing the transformed host cell produced in step (a) under conditions that are suitable for expression of the chimeric genewherein expression of the chimeric gene results in production of altered levels of a phytic acid biosynthetic enzyme in the transformed host cell.
7. A method of obtaining a nucleic acid fragment encoding all or a substantial portion of the amino acid sequence encoding a phytic acid biosynthetic enzyme comprising:
 - (a) probing a cDNA or genomic library with the nucleic acid fragment of Claim 1;
 - (b) identifying a DNA clone that hybridizes with the nucleic acid fragment of Claim 1;
 - (c) isolating the DNA clone identified in step (b); and

- (d) sequencing the cDNA or genomic fragment that comprises the clone isolated in step (c)

wherein the sequenced nucleic acid fragment encodes all or a substantial portion of the amino acid sequence encoding a phytic acid biosynthetic enzyme.

- 5 8. A method of obtaining a nucleic acid fragment encoding a substantial portion of an amino acid sequence encoding a phytic acid biosynthetic enzyme comprising:

- (a) synthesizing an oligonucleotide primer corresponding to a portion of the sequence set forth in any of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29 and 31; and

- 10 (b) amplifying a cDNA insert present in a cloning vector using the oligonucleotide primer of step (a) and a primer representing sequences of the cloning vector

wherein the amplified nucleic acid fragment encodes a substantial portion of an amino acid sequence encoding a phytic acid biosynthetic enzyme.

- 15 9. The product of the method of Claim 7.

10. The product of the method of Claim 8.

180
 SEQ ID NO:33 (gi 3396079) DW-----YKKNLHREFKCPGVPI-DLPEAIERLHNR-----VSMLEVTQL
 SEQ ID NO:34 (gi 3660465) DY-----YKKNLHREFKCPGVPI-DLPEAIERLHNR-----VSMLEVTQL
 SEQ ID NO:2 RK-----ENQVLXXEHEPEVTXL-DPEAIEHLNR-----ISMLEEVADL
 SEQ ID NO:4 GQ-----ENQVLESTINNEPNTV-DQFTSIQRHNR-----QSMLEVTQL
 SEQ ID NO:6 GD-----HEPNVTVL-DPPDAIQHNR-----QSMLOVDVL
 SEQ ID NO:8 GE-----DMRAQLVAFARQPVSS-TRPRH-RPLHNR-----ISMLOVSEL
 SEQ ID NO:10 GE-----HEPEVTVL-DPEGAIEHLNR-----QSMLOVSEL
 SEQ ID NO:12 GE-----EMRGQDAESAHPAPVY-DPEGAIEHLNR-----QSMLOVSEL
 SEQ ID NO:14 GD-----GMRQLEELAAHPEVY-DPEGAIEHLNR-----QSMLOVSEL
 SEQ ID NO:16 GD-----DMRQLOEERHRLHNR-----QSMLOVSEL
 SEQ ID NO:18 GE-----EXCEILEDTRKHPVTVL-DPEGAIEHLNR-----QSMLOVDVL
 SEQ ID NO:20 GE-----HWNLLQOESSKHPNTVII-DPELIVLRLHNR-----VSMLEVTQL
 SEQ ID NO:22 TD-----PHBQLEESALHPSVPVW-DAPAVORLLDRF-----TMLDVPGL
 SEQ ID NO:24 SEQ ID NO:26 SEQ ID NO:28 SEQ ID NO:30 SEQ ID NO:32
 181
 SEQ ID NO:33 (gi 3396079) REPVS--D-SEREGVPEQVVM-DSSV--LSGGGALGELKFEVIAKPLADGAGSKSHMF
 SEQ ID NO:34 (gi 3660465) REPVS--D-SEREGVPEQVVM-DSSV--LSGGGALGELKFEVIAKPLADGAGSKSHMF
 SEQ ID NO:2 NIP-----LSNFYGEVCTPRQLVTK--DFSSIPTSVAMGTLPLVARPLVVDGTSKSHLF
 SEQ ID NO:4 NIP-----LSDSYGTGVPRQLVIT--KNQPTSI DRAKAGLRLPWAKPLV-----AKSHELS
 SEQ ID NO:6 DHAVD--Q-DSTGTGTPQVYI--DAKGLDGLLAIRFELIAKPLVADGTAKSHMS
 SEQ ID NO:8 D--LSDCRGVTPVYI--NDPSSI PAAMVRAGLSLPLVAKPLV-----AKSHELS
 SEQ ID NO:10 D--LSDCRGVTPVYI--NDPSSI PAAMVRAGLSLPLVAKPLVADGTAKSHMS
 SEQ ID NO:12 D--LSDCRGVTPVYI--NDPSSI PAAMVRAGLSLPLVAKPLVADGTAKSHMS
 SEQ ID NO:14 D--LSDCRGVTPVYI--NDPSSI PAAMVRAGLSLPLVAKPLVADGTAKSHMS
 SEQ ID NO:16 RIE--D--B--PETGIPKQVILY-DKAT--LLDQWESLKFEVIAKPLVADGAGSKSHMA
 SEQ ID NO:18 NIP-----LSNFGVTPVYI--NDPSSI PAAMVRAGLSLPLVAKPLVADGTAKSHMA
 SEQ ID NO:20 NIP-----LSNFGVTPVYI--NDPSSI PAAMVRAGLSLPLVAKPLVADGTAKSHMA
 SEQ ID NO:22 QESLE--NATIGVPRQVVM-EPKSFDLHKFEEDQGLAFVIAKPLADGAGSHEL
 SEQ ID NO:24 AGLDFF-----LSVPAQVTKRRRAGRGF--STGSLEIAKPLAS--
 SEQ ID NO:26 D--LADCHCKGVPRQLV--NTDPLSI PAAMVRAGLSLPLVAKPLV-----AKSHELS
 SEQ ID NO:28 NIP-----LSNFGVTPVYI--NDPSSI PAAMVRAGLSLPLVAKPLVADGTAKSHMA
 SEQ ID NO:30 NIP-----LSNFGVTPVYI--NDPSSI PAAMVRAGLSLPLVAKPLVADGTAKSHMA
 SEQ ID NO:32 NIP-----LSNFGVTPVYI--NDPSSI PAAMVRAGLSLPLVAKPLVADGTAKSHMA

361
 SEQ ID NO:33 (g1 3396079)
 SEQ ID NO:34 (g1 3660465)
 SEQ ID NO:1
 SEQ ID NO:2
 SEQ ID NO:3
 SEQ ID NO:4
 SEQ ID NO:5
 SEQ ID NO:6
 SEQ ID NO:7
 SEQ ID NO:8
 SEQ ID NO:9
 SEQ ID NO:10
 SEQ ID NO:11
 SEQ ID NO:12
 SEQ ID NO:13
 SEQ ID NO:14
 SEQ ID NO:15
 SEQ ID NO:16
 SEQ ID NO:17
 SEQ ID NO:18
 SEQ ID NO:19
 SEQ ID NO:20
 SEQ ID NO:21
 SEQ ID NO:22
 SEQ ID NO:23
 SEQ ID NO:24
 SEQ ID NO:25
 SEQ ID NO:26
 SEQ ID NO:27
 SEQ ID NO:28
 SEQ ID NO:29
 SEQ ID NO:30
 SEQ ID NO:31
 SEQ ID NO:32

421
 SEQ ID NO:33 (g1 3396079)
 SEQ ID NO:34 (g1 3660465)
 SEQ ID NO:1
 SEQ ID NO:2
 SEQ ID NO:3
 SEQ ID NO:4
 SEQ ID NO:5
 SEQ ID NO:6
 SEQ ID NO:7
 SEQ ID NO:8
 SEQ ID NO:9
 SEQ ID NO:10
 SEQ ID NO:11
 SEQ ID NO:12
 SEQ ID NO:13
 SEQ ID NO:14
 SEQ ID NO:15
 SEQ ID NO:16
 SEQ ID NO:17
 SEQ ID NO:18
 SEQ ID NO:19
 SEQ ID NO:20
 SEQ ID NO:21
 SEQ ID NO:22
 SEQ ID NO:23
 SEQ ID NO:24
 SEQ ID NO:25
 SEQ ID NO:26
 SEQ ID NO:27
 SEQ ID NO:28
 SEQ ID NO:29
 SEQ ID NO:30
 SEQ ID NO:31
 SEQ ID NO:32

480
 SEQ ID NO:33 (g1 3396079)
 SEQ ID NO:34 (g1 3660465)
 SEQ ID NO:1
 SEQ ID NO:2
 SEQ ID NO:3
 SEQ ID NO:4
 SEQ ID NO:5
 SEQ ID NO:6
 SEQ ID NO:7
 SEQ ID NO:8
 SEQ ID NO:9
 SEQ ID NO:10
 SEQ ID NO:11
 SEQ ID NO:12
 SEQ ID NO:13
 SEQ ID NO:14
 SEQ ID NO:15
 SEQ ID NO:16
 SEQ ID NO:17
 SEQ ID NO:18
 SEQ ID NO:19
 SEQ ID NO:20
 SEQ ID NO:21
 SEQ ID NO:22
 SEQ ID NO:23
 SEQ ID NO:24
 SEQ ID NO:25
 SEQ ID NO:26
 SEQ ID NO:27
 SEQ ID NO:28
 SEQ ID NO:29
 SEQ ID NO:30
 SEQ ID NO:31
 SEQ ID NO:32

420
 SEQ ID NO:33 (g1 3396079)
 SEQ ID NO:34 (g1 3660465)
 SEQ ID NO:1
 SEQ ID NO:2
 SEQ ID NO:3
 SEQ ID NO:4
 SEQ ID NO:5
 SEQ ID NO:6
 SEQ ID NO:7
 SEQ ID NO:8
 SEQ ID NO:9
 SEQ ID NO:10
 SEQ ID NO:11
 SEQ ID NO:12
 SEQ ID NO:13
 SEQ ID NO:14
 SEQ ID NO:15
 SEQ ID NO:16
 SEQ ID NO:17
 SEQ ID NO:18
 SEQ ID NO:19
 SEQ ID NO:20
 SEQ ID NO:21
 SEQ ID NO:22
 SEQ ID NO:23
 SEQ ID NO:24
 SEQ ID NO:25
 SEQ ID NO:26
 SEQ ID NO:27
 SEQ ID NO:28
 SEQ ID NO:29
 SEQ ID NO:30
 SEQ ID NO:31
 SEQ ID NO:32

481 498
SEQ ID NO:33 (g1 3396079) V-----
SEQ ID NO:34 (g1 3660465) V-----
SEQ ID NO:2 L-----
SEQ ID NO:4 L-----
SEQ ID NO:6 L-----
SEQ ID NO:8 L-----
SEQ ID NO:10 L-----
SEQ ID NO:12 L-----
SEQ ID NO:14 L-----
SEQ ID NO:16 L-----
SEQ ID NO:18 L-----
SEQ ID NO:20 L-----
SEQ ID NO:22 L-----
SEQ ID NO:24 L-----
SEQ ID NO:26 L-----
SEQ ID NO:28 L-----
SEQ ID NO:30 L-----
SEQ ID NO:32 L-----

SEQUENCE LISTING

<110> E I. DU PONT DE NEMOURS AND COMPANY
 <120> PHYTIC ACID BIOSYNTHETIC ENZYMES
 <130> BB-1165-A
 <140>
 <141>
 <150> 60/082,960
 <151> APRIL 24, 1998
 <160> 34
 <170> Microsoft Office 97
 <210> 1
 <211> 1589
 <212> DNA
 <213> Zea mays
 <220>
 <221> unsure
 <222> (410)
 <220>
 <221> unsure
 <222> (478)
 <220>
 <221> unsure
 <222> (492)
 <220>
 <221> unsure
 <222> (519)
 <220>
 <221> unsure
 <222> (525)..(526)
 <220>
 <221> unsure
 <222> (534)
 <220>
 <221> unsure
 <222> (557)
 <400> 1
 ttattaaacag ctccgcgggtc cctccctccc tgggtcggtc ggcgtcggtc cctctccctc 60
 cccacccagtg tagtcctcag cctatcccggt gccgcgcag agcaccgct ctgctcgacc 120
 caccacccctc tgtgcagagt taattaacct tgagggttcc gattgcacct cctctccggt 180
 cctctcgccc attcgcggcg agattcagcg gcaaggatgc cctgcacgc ggaggtcgcg 240
 gatgagatgg aggaggggag cgaggtgggg gctgtgacgg cctcgcggg gctgtcgcca 300
 cgcacactca tgggtgcggc ggcgcgggtt ccccgatag tgggtgggtt cgcctcacg 360
 aagaagaagg tgaagagctt cctgcagccc aagctgctcc tgctgggcan gaagaatgga 420
 atcaggtttg tatctattga tgagtctctt cccctctcag aacaaggccc tttgatntr 480
 attttacaca anattactag gaaggagtgc cagaaggtnc tggannacta tcangaagaa 540
 catccagaag ttactgncct tgaccaccca aatgctatcg agcatctgaa caatcgacaa 600
 tcaatgcttg aagaagtgcg tgatttgaac ctgccaatt tctatggaga agtttctatc 660
 ccaagccagc tggctattac gaaagatcca tctctatac caactctgt agctatggct 720
 gacttaactt tgccttgggt tgccaagcca ttggttgttg atgggacgtc taaaggtcat 780
 gaactatctc ttgcatatga cgaggcatcc ttgcaatgc ttgatccgct tctggttcta 840

```

caggaattca taaacctgg cgggactctc tttaagggtg atactatttg tgaacaata 900
cagggtgtcc cgaggtttct tcttctctgat gtaacacat atgacttact aaacaacgtt 960
ggcatctatc gattgccaaag agtttcatgt gctgcagcta gtgcggatga tgcagatctt 1020
gacctctcta ttgcagagct tcttccaagg ccactctctag agaaaactggg caggggagctt 1080
cgtggcgctg ttggtttgag attgttcaat atagatatga ttagagaact tgggaaccaaa 1140
gatcggtact acataattga tatcaactac ttcccaggtt acggcaaaaat gccaggatat 1200
gagcgcatgt tcacagactt ctactaagt ctcgcacaaa gcaagtacaa aaggtactta 1260
agcggggcgt gagggtgcaag gaagtgtgtg aagaccatgc tactgacgag atggcatata 1320
acggtggcag ctatgtctcc ccaccgcgcc aatgtacatt tgcaggagac ataaacataa 1380
gccggaggtc tagggaagtt gycgaagtctc agtgtgtgtg ttcaaaaatcg gtggcacatg 1440
ctggactgga gtaggaaata accaaggaaa cgcttggtat cgctgtaccc atgtgtgaaa 1500
atgtttaact gaatgaacac cttctctgtg atggctctct ccactgtaat ttggcaacca 1560
tgagaattaa ttctgcaaaa aaaaaaaaaa 1589

```

```

<210> 2
<211> 351
<212> PRT
<213> Zea mays

<220>
<221> UNSURE
<222> (65)

<220>
<221> UNSURE
<222> (88)

<220>
<221> UNSURE
<222> (92)

<220>
<221> UNSURE
<222> (103)..(104)

<220>
<221> UNSURE
<222> (106)

<220>
<221> UNSURE
<222> (114)

<400> 2
Met Arg Leu His Ala Glu Val Arg Asp Glu Met Glu Glu Gly Ser Glu
1 5 10 15
Val Gly Ala Val Thr Ala Ser Ala Gly Leu Ser Pro Pro Pro Leu Ile
20 25 30
Gly Ala Ala Ala Pro Val Pro Arg Ile Val Val Gly Phe Ala Leu Thr
35 40 45
Lys Lys Lys Val Lys Ser Phe Leu Gln Pro Lys Leu Leu Leu Ala
50 55 60
Xaa Lys Asn Gly Ile Ser Phe Val Ser Ile Asp Glu Ser Leu Pro Leu
65 70 75 80
Ser Glu Gln Gly Pro Phe Asp Xaa Ile Leu His Xaa Ile Thr Arg Lys
85 90 95
Glu Trp Gln Lys Val Leu Xaa Xaa Tyr Xaa Glu Glu His Pro Glu Val
100 105 110

```

Thr Xaa Leu Asp Pro Pro Asn Ala Ile Glu His Leu Asn Asn Arg Gln
 115 120 125
 Ser Met Leu Glu Glu Val Ala Asp Leu Asn Leu Ser Asn Phe Tyr Gly
 130 135 140
 Glu Val Cys Ile Pro Arg Gln Leu Val Ile Thr Lys Asp Pro Ser Ser
 145 150 155 160
 Ile Pro Thr Ser Val Ala Met Ala Gly Leu Thr Leu Pro Leu Val Ala
 165 170 175
 Lys Pro Leu Val Val Asp Gly Thr Ser Lys Gly His Glu Leu Tyr Leu
 180 185 190
 Ala Tyr Asp Glu Ala Ser Leu Ser Met Leu Asp Pro Pro Leu Val Leu
 195 200 205
 Gln Glu Phe Ile Asn His Gly Gly Ile Leu Phe Lys Val Tyr Ile Ile
 210 215 220
 Gly Glu Thr Ile Gln Val Val Arg Arg Phe Ser Leu Pro Asp Val Asn
 225 230 235 240
 Thr Tyr Asp Leu Leu Asn Asn Val Gly Ile Tyr Arg Leu Pro Arg Val
 245 250 255
 Ser Cys Ala Ala Ala Ser Ala Asp Asp Ala Asp Leu Asp Pro Leu Ile
 260 265 270
 Ala Glu Leu Pro Pro Arg Pro Leu Glu Lys Leu Gly Arg Glu Leu
 275 280 285
 Arg Gly Arg Phe Gly Leu Arg Leu Phe Asn Ile Asp Met Ile Arg Glu
 290 295 300
 Leu Gly Thr Lys Asp Arg Tyr Tyr Ile Ile Asp Ile Asn Tyr Phe Pro
 305 310 315 320
 Gly Tyr Gly Lys Met Pro Gly Tyr Glu Arg Met Phe Thr Asp Phe Leu
 325 330 335
 Leu Ser Leu Ala Gln Ser Lys Tyr Lys Arg Tyr Leu Ser Gly Thr
 340 345 350

<210> 3
 <211> 494
 <212> DNA
 <213> Dimorphotheca sinuata

<220>
 <221> unsure
 <222> (7)

<220>
 <221> unsure
 <222> (179)

<220>
 <221> unsure
 <222> (392)

<220>
 <221> unsure
 <222> (405)

<220>
 <221> unsure
 <222> (442)

<220>
 <221> unsure
 <222> (461)

<220>
 <221> unsure
 <222> (476)

<220>
 <221> unsure
 <222> (489)

<220>
 <221> unsure
 <222> (493)

<400> 3
 attccancaa cacaaaaatg ccacaatttg tagaaaaaac cgatcttttt cgagtaggtt 60
 acgcttttacc taaaagaaga atcgaaagctt tcatgggtcga atcattcatc aactacgcta 120
 aagaacgaaa aatcgatttc atcccaatcg atgtttcaca accactaacc gaacaagcnc 180
 catccaatg cataattcac aagatgtatg gccagaatg gaaccaaaat ctcgaaagt 240
 tcactatcaa caaccctaac gccaccgtta tcgaccaacc aacctccatc caacgcctcc 300
 ataaccgaat ctgatgctt gaaccggtta cccaactcaa catcccaaaa ctcaacatac 360
 caaaccaact tttaggtcaa gatttcggtt gntttaattc ggtcncaca ataaaaggac 420
 ttaatttttc ctggtgatcg gngaaccgct attaagcggg ntggggaccac gaaggnacac 480
 gatagtctng ttng 494

<210> 4
 <211> 112
 <212> PRT
 <213> Dimorphotheca sinuata

<400> 4
 Met Pro Gln Phe Val Glu Lys Thr Asp Leu Phe Arg Val Gly Tyr Ala
 1 5 10 15
 Leu Pro Lys Arg Lys Ile Glu Ala Phe Met Val Glu Ser Phe Ile Asn
 20 25 30
 Tyr Ala Lys Glu Arg Lys Ile Asp Phe Ile Pro Ile Asp Val Ser Lys
 35 40 45
 Pro Leu Thr Glu Gln Gly Pro Phe Asn Cys Ile Ile His Lys Met Tyr
 50 55 60
 Gly Gln Glu Trp Asn Gln Asn Leu Glu Ser Phe Thr Ile Asn Asn Pro
 65 70 75 80
 Asn Ala Thr Val Ile Asp Gln Pro Thr Ser Ile Gln Arg Leu His Asn
 85 90 95
 Arg Ile Ser Met Leu Glu Pro Val Thr Gln Leu Asn Ile Pro Lys Leu
 100 105 110

<210> 5
 <211> 1136
 <212> DNA
 <213> Catalpa speciosa

<400> 5
 gcacgagcca aatgttacag ttcttgatcc tccagatgcc atacagcatg tccacaacgc 60
 tcaagtctatg cttcaggatg ttgctgaact gaattctgca gattcatatg gaacagtcgg 120

```

tgttctctaaa caattgggta taaaaaatga tccaacatct attcctgatg cagtgaacaa 180
ggctggggctg aggtctaccta tgggtggcaaa gccattgggt gctaagtcac atgagctgtc 240
tcttgccctat gatgagttct ccttcagaaa tcttgaaccc ccacttggtc tgcagggaatt 300
tattaatcac ggaggagtag tcttcaaagt ttattattgt ggggaagcaa taagggtggt 360
caggcgctttc tcttgccctg atgtgagtaa gcgtgaacta tcaagaatg ccgggtgtta 420
tcgcttttcca agagtatctt gtgcggtgc atctgctgat gaagcagact tggatccttg 480
tgtagctgag ctcctctctc gccctactact tgagaaactg gctaggaatc tacgtcatcg 540
gctgggtctt cgactattca acttggatgt gattcgtgag cacggaactc gagaccacta 600
ctatgttata gatattaact actttccnag atacgggaaa atgcagcaat atgagcatat 660
attcacgat ttctctttga gccttggttca aagcaaatag aaaagtgtaa tcccgcaag 720
agagtgccgag aaggatttga gtgcccccat gctcgacaag aagtgtact ccaaaacaaa 780
atcacgcaac aggaagaagt gaagccagtc ctgtgataag tttttgatga ttattgaagg 840
acaaagaatg caaaagatcg attcgtttgt tctcatctct ggcaaaaatt gtcattttaga 900
aggtcctctt ggaagaaggt tcaaattatc ctctgggaaa ggcggttttt ttgtttcata 960
ttttttgatt ttcgtttcat ttttctcatt ttttttgctg gttttcaagt tggattgctc 1020
ttgatataata ttttgacatt gttagtact agttgtgcc agtcattct atatttacgg 1080
cctttttgtt ctcacaaaaa aaaaaaaaaa aactcgaggg gggggccgta cacaat 1136

```

<210> 6
 <211> 232
 <212> PRT
 <213> Catalpa speciosa

```

<400> 6
His Glu Pro Asn Val Thr Val Leu Asp Pro Pro Asp Ala Ile Gln His
 1 5 10 15

Val His Asn Arg Gln Ser Met Leu Gln Asp Val Ala Asp Leu Asn Leu
 20 25 30

Ser Asp Ser Tyr Gly Thr Val Gly Val Pro Lys Gln Leu Val Ile Lys
 35 40 45

Asn Asp Pro Thr Ser Ile Pro Asp Ala Val Asn Lys Ala Gly Leu Arg
 50 55 60

Leu Pro Met Val Ala Lys Pro Leu Val Ala Lys Ser His Glu Leu Ser
 65 70 75 80

Leu Ala Tyr Asp Glu Phe Ser Leu Gln Asn Leu Glu Pro Pro Leu Val
 85 90 95

Leu Gln Glu Phe Ile Asn His Gly Gly Val Leu Phe Lys Val Tyr Ile
 100 105 110

Val Gly Glu Ala Ile Lys Val Val Arg Arg Phe Ser Leu Pro Asp Val
 115 120 125

Ser Lys Arg Glu Leu Ser Lys Asn Ala Gly Val Tyr Arg Phe Pro Arg
 130 135 140

Val Ser Cys Ala Ala Ser Ala Asp Glu Ala Asp Leu Asp Pro Cys
 145 150 155 160

Val Ala Glu Leu Pro Pro Arg Pro Leu Leu Gly Lys Leu Ala Arg Asn
 165 170 175

Leu Arg His Arg Leu Gly Leu Arg Leu Phe Asn Leu Asp Val Ile Arg
 180 185 190

Glu His Gly Thr Arg Asp His Tyr Tyr Val Ile Asp Ile Asn Tyr Phe
 195 200 205

Pro Gly Tyr Gly Lys Met Pro Glu Tyr Glu His Ile Phe Thr Asp Phe
 210 215 220

```

Leu Leu Ser Leu Val Gln Ser Lys
225 230

<210> 7
<211> 1197
<212> DNA
<213> Zea mays

<220>
<221> unsure
<222> (311)

<220>
<221> unsure
<222> (330)

<220>
<221> unsure
<222> (849) .. (850)

<220>
<221> unsure
<222> (941)

<220>
<221> unsure
<222> (1112)

<220>
<221> unsure
<222> (1138)

<220>
<221> unsure
<222> (1156)

<220>
<221> unsure
<222> (1190)

<400> 7
gcaaatattca atctccatcg atcgattcct cccgaaccgg acccgatggc ctccgacggc 60
gccgcgagcg cctctctccgg cgtcaccac ccccgcgct acgtcatcg ttacgcgctc 120
gccgcgaaaga agcagcaaaag ctctatccag cgtctcgctg tggcccaagg gccgctcgcg 180
ggcatcgagcc tcgtcccgct ggatgcgctg cagccctgg cagagcaagg gcccttccac 240
ctctctatcc acaagctcta cggagacgac tggcgcgccc agctcgtggc ctccgcccgg 300
cgccaaccgg ncgtcccatc gtcgaccgcn caacgccat gaccgctcca caacggcatc 360
tccatgctcc aggtcgtctc cgagctcgac caccgctcg accaggacag cacttctgg 420
atccccagcc aggtcgtcgt ctacgacgct gccgcgctcg ccgacttcgg actcttggc 480
ggcgtccgct tcccgcctcat cggcgaagccc ctcgctcgcg acggcaaccg caagtccac 540
aagatgttcg tcgtctacca cggcgaggcg ctccgcaagg tccgcccgc gcttgtgctc 600
caggagttcg tcgtctacca cggcgaggcg ctccgcaagg tccgcccgc gcttgtgctc 660
acttgcgca agcgcgtag cctgcccagc gtgtcccgc aggatgacg atcgcccagc 720
ggatcgtct ccttctccca ggtctccaac ctcccactg agcgcacggc ggaggagtac 780
tacggcgaaa agatgtctga ggacgcgctg gtgcggccg ccgcatctat caaccagatc 840
gcggcggnn ctccgcgccc gctgggctg caactcttca acttcgacat gatccgcgac 900
ctccgcgccc gcgacgcta tctcgtcatt gacatcaacc ntacttcggg ctacgccaag 960
atgccaggat acgagactgt ctacggattt ctctctggag atggtccata aggacggcgt 1020
gggcaaccaa caggaggaag aaaggggcca accattgttg tcgggaaata agaatgatga 1080
attgatggca acttgggata attcttgggc cnaaaatggc ttggcttggg aattccctng 1140
gaatggccaa gaaaanttcc gaattggaaa gggggggaaa tttttaaana ttggggg 1197

<210> 8
<211> 321
<212> PRT
<213> Zea mays

<220>
 <221> UNSURE
 <222> (89)

<220>
 <221> UNSURE
 <222> (269)

<220>
 <221> UNSURE
 <222> (299)

<400> 8
 Met Ala Ser Asp Ala Ala Ala Glu Pro Ser Ser Gly Val Thr His Pro
 1 5 10 15
 Pro Arg Tyr Val Ile Gly Tyr Ala Leu Ala Pro Lys Lys Gln Gln Ser
 20 25 30
 Phe Ile Gln Pro Ser Leu Val Ala Gln Ala Ala Ser Arg Gly Met Asp
 35 40 45
 Leu Val Pro Val Asp Ala Ser Gln Pro Leu Ala Glu Gln Gly Pro Phe
 50 55 60
 His Leu Leu Ile His Lys Leu Tyr Gly Asp Asp Trp Arg Ala Gln Leu
 65 70 75 80
 Val Ala Phe Ala Ala Arg Gln Pro Xaa Val Pro Ser Ser Thr Arg Pro
 85 90 95
 Arg His Arg Pro Leu His Asn Arg Ile Ser Met Leu Gln Val Ser
 100 105 110
 Glu Leu Asp His Ala Val Asp Gln Asp Ser Thr Phe Gly Ile Pro Ser
 115 120 125
 Gln Val Val Val Tyr Asp Ala Ala Ala Leu Ala Asp Phe Gly Leu Leu
 130 135 140
 Ala Ala Leu Arg Phe Pro Leu Ile Ala Lys Pro Leu Val Ala Asp Gly
 145 150 155 160
 Thr Ala Lys Ser His Lys Met Ser Leu Val Tyr His Arg Glu Gly Leu
 165 170 175
 Gly Lys Leu Arg Pro Pro Leu Val Leu Gln Glu Phe Val Asn His Gly
 180 185 190
 Gly Val Ile Phe Lys Val Tyr Val Val Gly Gly His Val Thr Cys Val
 195 200 205
 Lys Arg Arg Ser Leu Pro Asp Val Ser Pro Glu Asp Asp Ala Ser Ala
 210 215 220
 Gln Gly Ser Val Ser Phe Ser Gln Val Ser Asn Leu Pro Thr Glu Arg
 225 230 235 240
 Thr Ala Glu Glu Tyr Tyr Gly Glu Lys Ser Leu Glu Asp Ala Val Val
 245 250 255
 Pro Pro Ala Ala Phe Ile Asn Gln Ile Ala Gly Gly Xaa Pro Arg Ala
 260 265 270

Leu Gly Leu Gln Leu Phe Asn Phe Asp Met Ile Arg Asp Val Arg Ala
 275 280 285

Gly Asp Arg Tyr Leu Val Ile Asp Ile Asn Xaa Thr Ser Gly Tyr Ala
 290 295 300

Lys Met Pro Gly Tyr Glu Thr Val Tyr Gly Phe Leu Leu Gly Asp Gly
 305 310 315 320

Pro

<210> 9
 <211> 518
 <212> DNA
 <213> Oryza sativa

<220>
 <221> unsure
 <222> (366)

<220>
 <221> unsure
 <222> (391)

<220>
 <221> unsure
 <222> (431)

<220>
 <221> unsure
 <222> (438)

<220>
 <221> unsure
 <222> (447)

<220>
 <221> unsure
 <222> (463)

<220>
 <221> unsure
 <222> (500)

<220>
 <221> unsure
 <222> (509)

<400> 9
 ggaacaccca gaagttactg ttcttgaccc accaggtgcc atagaacatc tgccttaacg 60
 tcaatctatg ctccaagagg ttcttgaatt ggaacctcca gattgtcatg gtgaggtcgg 120
 tgttccaaag cagctgtttg ttaatacaga tccatcacca ataccagctg ctgttatgag 180
 ggctggctca tctcttccat tagtggcaca gcctctgggt gcaaatccc atgagttatc 240
 tcttggctta gctccaatct ccttgacgaa gcttgagcct cgcctgtgtc tcaggaatt 300
 tgaacaacat ggtggtgtct tgtttaaggt ctacattgtt ggggatgcca tacgggttgt 360
 gcgtangttt tgccttccca acgttgatgt nggtgatatta tcaacaacat ctggagttat 420
 ccggtttcaa nggtctctntg tgctcangca aaccagatat ganattgggc cccatgttct 480
 gaactcccca agacactccn agatattgna agagagtg 518

<210> 10
 <211> 244
 <212> PRT
 <213> Oryza sativa

<400> 10
 Glu His Pro Glu Val Thr Val Leu Asp Pro Pro Gly Ala Ile Glu His
 1 5 10 15
 Leu Leu Asn Arg Gln Ser Met Leu Gln Glu Val Ser Glu Leu Asp Leu
 20 25 30
 Ser Asp Cys His Gly Arg Val Gly Val Pro Lys Gln Leu Phe Val Asn
 35 40 45
 Thr Asp Pro Ser Ser Ile Pro Ala Ala Val Met Arg Ala Gly Leu Ser
 50 55 60
 Leu Pro Leu Val Ala Lys Pro Leu Val Ala Lys Ser His Glu Leu Ser
 65 70 75 80
 Leu Ala Tyr Asp Pro Ile Ser Leu Thr Lys Leu Glu Pro Pro Leu Val
 85 90 95
 Leu Gln Glu Phe Val Asn His Gly Gly Val Leu Phe Lys Val Tyr Ile
 100 105 110
 Val Gly Asp Ala Ile Arg Val Val Arg Arg Phe Ser Leu Pro Asn Val
 115 120 125
 Asp Val Gly Asp Leu Ser Asn Asn Ala Gly Val Phe Arg Phe Pro Arg
 130 135 140
 Val Ser Cys Ala Ser Ala Asn Ala Asp Asp Ala Asp Leu Asp Pro His
 145 150 155 160
 Val Ala Glu Leu Pro Pro Arg Pro Leu Leu Glu Ile Leu Ala Arg Glu
 165 170 175
 Leu Arg Arg Arg Leu Gly Leu Arg Leu Phe Asn Ile Asp Met Ile Arg
 180 185 190
 Glu His Gly Thr Arg Asp Arg Phe Tyr Val Ile Asp Met Asn Tyr Phe
 195 200 205
 Pro Gly Tyr Gly Lys Met Pro Gly Tyr Glu His Val Phe Thr Asp Phe
 210 215 220
 Leu Leu Ser Leu Val Gln Lys Glu Tyr Lys Arg Arg Pro Ser Tyr Ser
 225 230 235 240
 Ser Cys Glu Gly

<210> 11
 <211> 1574
 <212> DNA
 <213> Oryza sativa

<400> 11
 gcacgaggaa gagggaaatc gcgtgggcga ggaggaggga ggaagagagag ccgcttccgc 60
 ttgatttcgat tcgaagcaaa accaaccctaa cccaaccctaa ctcaaccctaa tcgatggccg 120
 gcgacgagcc cctccccgcg gcgaggcagc gcgcgcgcga cctcatcgcc tacgcctctg 180
 gcgcgagagaa gcagcagagc ttcatccagc cgtcgtcgtt gtcgcgcgcg gcggggcgcg 240
 gcatggacct cgtccccgtc gaccctctgc ggcctctccc cgagcagggg cctttccacc 300
 tctctatcca caagctctac ggcgaggagt ggcgcggtca gctggagccc tcttccgcgc 360
 cgacccagcg cgtccccgtc gtcgaccgcg cccacgccat gcacgcctc cacaaccgca 420
 tctcctatgt ccaggtcgtc tccgagctcg acgtcccgct ccacgcccac caccaccaca 480
 ccttcgcat cccctccag gtcgtcgtc acgacgcgcg gcctctcttc gactccggcc 540
 tctcgcgcg cctccgcttc cccctcatcg ccaagccct cgtcgcgcg ggaaccgca 600
 agtcccacaa gatgtccctc gtctaccacc gcgaggcgct ccgaagctc gcgccccgcg 660

tgcgtctcca ggagttcgtc aaccacggcg gctgcattct caaggtctac gtcgtcggcg 720
 cccacgtcac ctgcgtcaag cccgcagccc tcccgcagct ctccagcgac gtctctcagg 780
 acgcctccgg cgagggtccc ctctccttct cccaggtctc caacctcccc aacgaggcgc 840
 acgcccagga gtactacgac gacatgcgcc tcgaggacgc catcatgccca cccacgcgct 900
 tcacaagaaa cttggccgccc ggccctcgcc gctggggcct tccactcttc aaatttgaca 960
 tgatccggga ccccccgct gggaatcgct acctcgtcat tgacatcaat tacttccccg 1020
 gctatgccaa gatgcccgga tacgagactg tgctcacgga ttctctctgg gagatgggtc 1080
 acagagcaga cgacacgccc aacctcaacc ccaaccccaa tgacgaggat gtcaaatgag 1140
 ccaacatacat ttgcactgcc atgctcgcaa acgtgcaatg cctggatgct agcgcgaggt 1200
 accaggtatg ttgcactgcc atgctcgcaa acgtgcaatg aattaggatg 1260
 gggagatcgg ctggttcgag aaatctcgat gaggggaata acgagcaagc aattaggatg 1320
 cctctactcgc ggtgtaaatc ctcttctcta actcctttca tactggttgt tttaggatc 1380
 ctctcgttct tcgtaatttg ctctcatctg tttagttaga gcagccaaga gagagaaaag 1440
 aaatgaagaa cgaagtcttc aaagaaaatt tgctcgggtc aaaaaaaaaa aaaaaaaaaa 1500
 aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa tcgagggggg 1560
 gcccggtaca catt 1574

<210> 12
 <211> 341
 <212> PRT
 <213> Oryza sativa

<400> 12
 Met Ala Gly Asp Glu Pro Leu Pro Gly Asp Gly Gln Arg Arg Tyr
 1 5 10 15
 Leu Ile Gly Tyr Ala Leu Ala Pro Lys Lys Gln Gln Ser Phe Ile Gln
 20 25 30
 Pro Ser Leu Val Ser Arg Ala Ala Gly Arg Gly Met Asp Leu Val Pro
 35 40 45
 Val Asp Pro Ser Arg Pro Leu Pro Glu Gln Gly Pro Phe His Leu Leu
 50 55 60
 Ile His Lys Leu Tyr Gly Glu Glu Trp Arg Gly Gln Leu Asp Ala Phe
 65 70 75 80
 Ser Ala Ala His Pro Ala Val Pro Val Val Asp Pro Pro His Ala Ile
 85 90 95
 Asp Arg Leu His Asn Arg Ile Ser Met Leu Gln Val Val Ser Glu Leu
 100 105 110
 Asp Val Pro Leu His Ala His His His Thr Phe Gly Ile Pro Ser
 115 120 125
 Gln Val Val Val Tyr Asp Ala Ala Leu Ser Asp Ser Gly Leu Leu
 130 135 140
 Ala Ala Leu Arg Phe Pro Leu Ile Ala Lys Pro Leu Val Ala Asp Gly
 145 150 155 160
 Thr Ala Lys Ser His Lys Met Ser Leu Val Tyr His Arg Glu Gly Leu
 165 170 175
 Arg Lys Leu Arg Pro Pro Leu Val Leu Gln Glu Phe Val Asn His Gly
 180 185 190
 Gly Val Ile Phe Lys Val Tyr Val Val Gly Ala His Val Thr Cys Val
 195 200 205
 Lys Arg Arg Ser Leu Pro Asp Val Ser Ser Asp Val Leu Gln Asp Ala
 210 215 220

Ser Gly Glu Gly Ser Leu Ser Phe Ser Gln Val Ser Asn Leu Pro Asn
 225 230 235 240
 Glu Ala His Ala Gln Glu Tyr Tyr Asp Asp Met Arg Leu Glu Asp Ala
 245 250 255
 Ile Met Pro Pro Thr Ala Phe Thr Lys Asn Leu Ala Ala Gly Leu Ala
 260 265 270
 Arg Leu Gly Leu Pro Leu Phe Lys Phe Asp Met Ile Arg Asp Pro Pro
 275 280 285
 Ala Gly Asn Arg Tyr Leu Val Ile Asp Ile Asn Tyr Phe Pro Gly Tyr
 290 295 300
 Ala Lys Met Pro Gly Tyr Glu Thr Val Leu Thr Asp Phe Phe Trp Glu
 305 310 315 320
 Met Val His Lys Asp Asp Thr Pro Asn Leu Asn Pro Asn Pro Asn
 325 330 335
 Asp Glu Asp Val Lys
 340

<210> 13
 <211> 511
 <212> DNA
 <213> *Oryza sativa*

<400> 13
 cagctgggtca ctaccaacaa gtacaacaca caagtcacaa catatatata cctcgcatcg 60
 catcgattcc atcgattcgc gcggccatgg cgcccgagct gtccctcccg tcgtcgtcac 120
 ctccgtacac cgttggtcac gcctcgtcgc cgagaaggt gacgagcgtc gtgcggccgt 180
 cgtcgttggc gctggcgccc gaccgcgggg tgcgctcgt cgccgtcgac gtgtcgcggc 240
 cgctgcggca gacgggccgc ttgcacctcc tcgtgcacaa gatgtacgac cgcgggtggc 300
 gcgcccagct gaggagctc gcgcgcgcgc accccggggt gaccgtcgtc tcgactcc 360
 ccggcgccat cgaccgcctc ctcgaccgcgc ccaccatgct cgacgtcgtc tccgggtccc 420
 gcaacccggt ctccgtccgg ccccgagtcg tcgtcagcga cgccgcgcgc acgcggagca 480
 gtcctcctgc cgcgcgcgt ccgcttctcg g 511

<210> 14
 <211> 141
 <212> PRT
 <213> *Oryza sativa*

<400> 14
 Met Ala Pro Glu Leu Ser Ser Pro Ser Ser Ser Pro Arg Tyr Thr Val
 1 5 10 15
 Gly Tyr Ala Leu Leu Pro Glu Lys Val Ser Ser Val Val Arg Pro Ser
 20 25 30
 Leu Val Ala Leu Ala Ala Asp Arg Gly Val Arg Leu Val Ala Val Asp
 35 40 45
 Val Ser Arg Pro Leu Ala Glu Gln Gly Pro Phe Asp Leu Leu Val His
 50 55 60
 Lys Met Tyr Asp Arg Gly Trp Arg Ala Gln Leu Glu Glu Leu Ala Ala
 65 70 75 80
 Arg His Pro Gly Val Thr Val Val Val Asp Ser Pro Gly Ala Ile Asp
 85 90 95
 Arg Leu Leu Asp Arg Ala Thr Met Leu Asp Val Val Ser Gly Leu Arg
 100 105 110

Asn Pro Val Ser Val Arg Pro Gln Val Val Val Ser Asp Ala Ala Ala
115 120 125

Thr Arg Thr Ser Ser Leu Ala Ala Arg Ala Pro Leu Leu
130 135 140

<210> 15
<211> 1802
<212> DNA
<213> Glycine max

<400> 15
aacgcaatta atgtgagtta gctcactcat taggcacccc aggcctttaca ctttatgctt 60
cgcgctcgta tgttgtgtgg aattgtgagc ggataacaat ttcacacagc aaacagctat 120
gaccatgatt acgccaagcg cgcaattaaac cctcataaaa ggggaacaaa gctggagctc 180
acacccgggtg cgggcgcgctc tagaactagt ggaatccccg ggctgcagga attcggcagc 240
agccctagtc taagctatat atatccccac tccgttccct attcccttc tgcctctcct 300
ttttctctcc cctcactcct cggtccggtg gttcagaata ttcatcacca tcgggcccag 360
caccaaaaccc ctctccctct ccccaaaat agggttctc tctgggaagt tgatccagc 420
cgcgcgagca tcaatgggg agaagagatt cggcgtgata ggttaagctc tggcgcgcaa 480
gactcgtgcg gtggactccg acaagccctc cgcgatcac gcccctctc actcgcgtct 540
ccacaagctc tacggcgagc actggaagcg ccagctccag gaattccaca cctctcacc 600
aaaagcccgct atcctcgagc ccccgaggc gatcgagcg ctcacaacc gaatctcgat 720
gctgcagctc gtgtcggagc tccgaattga ggaccgacc gagacctttg ggaatcccaa 780
gcagattgtg atatacgaca aggccaccct cctggaccgc caggcctggg agagcctcaa 840
gtttcccggt atagcggaag ctctcgttgc cgacggcagc gccaaagtgc acaagatggc 900
gctcgtgttc acgcgcgagc cactgaacaa gctcaagccg ccgatcgtgc tgcaggaagt 960
cgtgaaccac ggtggcgtga tcttcaaggt gtacgtgggt ggcgagcag tgcgtctcgt 1020
gaagcgcaag tccctcccg acgtgcctga cgaggaaga ggcctcgcg ggtttccga 1080
ggactcgtat gtgtctcgc aggtctccaa cctcgcaacc gtgaacgat gcgacggcta 1140
ttacggtttg atgctctcgc acgacgacac cgaatggccc cccgacgctc tctgctcga 1200
catcgctggg ggctcagc gcgcctcaca gctcaacctc ttcaatttc atgtcataag 1260
agagcgtcga tacgggaacc gttatcttat cattgataac aactatttcc ctgggtacgc 1320
caaaatgccg ggctacgagc ccgttttgac ccagtttttc tgcgaagtaa tgcctcaaga 1380
gaaacagcaa gaggaacaa agcaagaaga ggttaacgct ctaaggaaga aagaggaaat 1440
tcttcaagct tgagataatt ttgtcacaca gatgtgttgc tgcggtgct ggtgtttggt 1500
ggtttttggg taacccgaat ttccccagtg gaacctatgg ttgtgtgctc ttgcaaat 1560
cgatttaggg tgcaattggg acagtagggg tgtttgtgct tagtttttt taacctctt 1620
caacaacatg tatattcctt tcaacctcgc tctctggaga taggtacct ttcaaatacc 1680
ctcaacattt aggaaggtg tattttgcgg ctaagatta ccaagatga aactgttac 1740
ttgtgttttt tttaactcaa taactcatt ttccattt tgccaaaaaa aaaaaaaa 1802
aa

<210> 16
<211> 399
<212> PRT
<213> Glycine max

<400> 16
Ala Ile Tyr Ile Pro Thr Pro Phe Pro Ile Ser Leu Leu Ser Phe Leu
1 5 10 15
Phe Leu Ser Leu Thr His Pro Phe Arg Trp Phe Arg Ile Phe Ile Thr
20 25 30
Ile Gly Pro Ser Thr Lys Pro Leu Pro Leu Ser Pro Lys Leu Gly Phe
35 40 45
Leu Ser Gly Asn Val Ile Gln Arg Ala Ala Ser Met Ala Glu Lys
50 55 60
Arg Phe Gly Val Ile Gly Tyr Ala Leu Ala Pro Lys Lys Gln Asn Ser
65 70 75 80

<400> 17
 ctctctctct tcccacaaat cgtgacacat ttctctcttt cctctctcta ttccgcgcga 60
 ttccgcaaat cccagattga gccccgccgc cgcgcaaat gaagaagaga aggcacaaag 120
 atttgatctt tcaatcgagg ggtttctggg gtgtccgagg atgaggctca acggtgaaat 180
 ctcaagtgga gaagaagagg agaaacaaac gggaaacgag acgtttctgt cgacagaagt 240
 ggtcgttgcc tacgctttaa cgtccaaaga gaaaaagagc ttctgcagc caagtttccac 300
 tggcctcgca cggaaacagg ggataaaact tgttgctatt gatctaaaca agccacgct 360
 tattgagga tataggcaaa aacatccaga ggttaactgt cttgatctc cggatgcaat 420
 ccaacattta cacaatcgcc aatccatgca gctagtatt ccaaaagaga aagacccttc 480
 ttgtcatggc aaggttggtg ttccagctga gctagtatt ccaaaagaga aagacccttc 540
 tagtatccct taacgaatca ctaaggctgg gatgaattg ccattagtgt caaaaccact 600
 agtgtggat ggcactgcaa agtcacatga actatttctt gcttatgatg aattctctct 660
 ttccagagct gaacctccac tagttctaca agagtgtgt aatcatggtg gttctctctt 720
 taagatttac attgttggtg aaaccataaa ggttgtgaag cgtttctctc ttcttaacat 780
 cagtaagcat aggttatcaa aagttgctgg tgtattccgt ttccaagag ttctatgtgc 840
 agccgcttct gcagatgatg ctgattttag tctaatatc gctgaacatc cgccaagacc 900
 ttattggag aggttgctga gggagctccg ccacgatgg ggactgtgct tttcactact 960
 agatatgatt cgaaatatg gaaccaagga tgtgtttac gtcattgaca tcaactactt 1020
 tcccggatg gaaaaaatgc cagactatga gcacgtattt acagattttc tactgaagct 1080
 agtcagagc aattgtaaga agaaacttgc taccctaaact agagcaagtt ttatgagct 1140
 agatccattg agcagggtcg ttctgtctgg tggcttctt aataatttat gcaaaactact 1200
 ctggaatttt aatgaagcca ccaaaatggt gaaaaaaaaa aaaaaa 1260
 1306

<210> 18
 <211> 338
 <212> PRT
 <213> Glycine max

<400> 18
 Met Arg Leu Asn Gly Glu Ile Ser Ser Gly Glu Glu Glu Glu Lys Gln 15
 1 5 10
 Thr Gly Thr Thr Phe Ser Ser Gln Lys Val Val Val Gly Tyr Ala 30
 20 25
 Leu Thr Ser Lys Lys Lys Lys Ser Phe Leu Gln Pro Ser Phe Thr Gly 45
 35 40
 Leu Ala Arg Asn Arg Gly Ile Asn Phe Val Ala Ile Asp Leu Asn Lys 60
 50 55
 Pro Leu Leu Glu Gln Gly Pro Phe Asp Ile Ile Leu His Lys Leu Ser 80
 65 70
 Gly Glu Glu Trp Cys Glu Ile Ile Glu Asp Tyr Arg Gln Lys His Pro 95
 85 90
 Glu Val Thr Val Leu Asp Pro Pro Asp Ala Ile Gln His Leu His Asn 110
 100 105
 Arg Gln Ser Met Leu Gln Asp Val Val Asp Leu Asn Leu Ser Asp Cys 125
 115 120
 His Gly Lys Val Gly Val Pro Arg Gln Leu Val Ile Pro Lys Glu Lys 140
 130 135
 Asp Pro Ser Ser Ile Pro Tyr Glu Ile Thr Lys Ala Gly Met Lys Leu 160
 145 150 155
 Pro Leu Val Ala Lys Pro Leu Val Val Asp Gly Thr Ala Lys Ser His 175
 165 170
 Glu Leu Phe Leu Ala Tyr Asp Glu Phe Ser Leu Ser Glu Leu Glu Pro 190
 180 185 190

Pro Leu Val Leu Gln Glu Phe Val Asn His Gly Gly Leu Leu Phe Lys
 195 200 205

Ile Tyr Ile Val Gly Glu Thr Ile Lys Val Val Lys Arg Phe Ser Leu
 210 215 220

Pro Asn Ile Ser Lys His Glu Val Ser Lys Val Ala Gly Val Phe Arg
 225 230 235 240

Phe Pro Arg Val Ser Cys Ala Ala Ser Ala Asp Asp Ala Asp Leu
 245 250 255

Asp Pro Asn Ile Ala Glu His Pro Pro Arg Pro Leu Leu Glu Arg Leu
 260 265 270

Ala Arg Glu Leu Arg His Arg Leu Gly Leu Cys Leu Phe Asn Ile Asp
 275 280 285

Met Ile Arg Glu Tyr Gly Thr Lys Asp Val Phe Tyr Val Ile Asp Ile
 290 295 300

Asn Tyr Phe Pro Gly Tyr Gly Lys Met Pro Asp Tyr Glu His Val Phe
 305 310 315 320

Thr Asp Phe Leu Leu Ser Leu Val Gln Ser Asn Cys Lys Lys Lys Leu
 325 330 335

Ala Thr

<210> 19
 <211> 910
 <212> DNA
 <213> Glycine max

<400> 19
 ttcttttcaa ggtttatata gttggtgatg ctataaagggt cgttaggcgg ttttcattac 60
 ctgatgtaag caagtgggaa ctctcgaaag atgctgggat atatcgtttt ccaagggttt 120
 cttgtgctgc agcttctgca gatgatgctg aattggatcc tactgttgcg gagcttcctc 180
 caagacccttt actagagaaa ctggctaagg aacttcgatg gcgatgggtg ctctgcttat 240
 tcaaacctgga tattatccgc gaglatggaa caagaatatca cttttaagtc attgacataa 300
 actactctcc tggatatggc aaatgccaag aatatgaaca tataattaca gacttctctg 360
 tgaagcttggg cgaagggaag tacaagaaaa aataggctaa aataactggt tccagccctt 420
 ttatttttgt cccccacatg acgtaagccc ccttgccctc cagtttgaag ctcaaaacta 480
 tggctgtagg aaagtattctt agaaagcatt agtgataata acttttccag ttatgctgtg 540
 tgatgggaagt gccatggaaa actatcaaga cttgtgttcc aggaacttcg acctgtacag 600
 aacaactgct agtattttat tgtgcgggaa ctcttcgctg tccttccaag gtaacaaatg 660
 ctgttgcaaa tcctctgaa gttctataagc tggtagcttg atgaacgaag gggctcaagt 720
 tgcaatttga aggatttgtc ttcgcaccgc agtggatgtt aatatgaccg tgttgataa 780
 gtgacgggat ttcaaatgta ctgacccggt ctttgctgta tataaggcat aaattcttct 840
 cgcctcgtgc cttttgcgct ggtatggata gtcaatgtgt tctgagcgcg cacggtgacg 900
 atagaagtgt

<210> 20
 <211> 130
 <212> PRT
 <213> Glycine max

<400> 20
 Leu Phe Lys Val Tyr Ile Val Gly Asp Ala Ile Lys Val Val Arg Arg
 1 5 10 15

Phe Ser Leu Pro Asp Val Ser Lys Trp Glu Leu Ser Lys Asp Ala Gly
 20 25 30

Ile Tyr Arg Phe Pro Arg Val Ser Cys Ala Ala Ala Ser Ala Asp Asp
 35 40 45
 Ala Asp Leu Asp Ser Pro Thr Val Ala Glu Leu Pro Pro Arg Pro Leu Leu
 50 55 60
 Glu Lys Leu Ala Lys Glu Leu Arg Trp Arg Leu Gly Leu Arg Leu Phe
 65 70 75 80
 Asn Leu Asp Ile Ile Arg Glu Tyr Gly Thr Arg Asn His Phe Tyr Val
 85 90 95
 Ile Asp Ile Asn Tyr Phe Pro Gly Tyr Gly Lys Met Pro Glu Tyr Glu
 100 105 110
 His Ile Phe Thr Asp Phe Leu Leu Ser Leu Gly Gln Gly Lys Tyr Lys
 115 120 125
 Lys Lys
 130
 <210> 21
 <211> 1040
 <212> DNA
 <213> Glycine max
 <400> 21
 gcaccaggca ggtcagaggt accgtatggg ctatgtcttc caaggcaaga aagtcgaaag 60
 ctctcatcaa cctcactcc tcgatcacgc caaacaacac agcatcgatc tcgtccaaat 120
 cgaccccaacc gcacctctac aacaacaagg tcctttccac tgcatacttc acaaactcca 180
 caccacaacac tggaaaaaac tcctccaaca attctcatcc aaacacccaa acacccgtaat 240
 catcgacctt ccgagctggg tggatgcgct gcacaaccgg gttccaatgc tcgacgcaat 300
 gaaccaactta caattttccc tcgaaaacgc caccattggg gttccaaggc aagtggttgc 360
 gaacgaaccc aaatccttgc atttgacaaa attcgaagaa gaacagggct tgcggttccc 420
 ggtgattgag aaaccgctgg cggctgacgg cggcgccggc tctcacgaac tgtgtttggt 480
 ttctgacgag gagggactcc acgcgttgag cgttcccatg gtgtgcgaag agttcgtgaa 540
 tcacggcggg gtcgtgttca agattttacgt tgcgtggcag cgcgtgaatt gcgtaaacgc 600
 caagtctcttg ggtgacataa cggaaagaaa gctgaaagtg ttaagggggg cgctgcgctt 660
 ttctcgtgtg tcgagtttgg ggggtgaaga cgaagggtgg ggcgcggttg aggcagctga 720
 aatgcctccg cagagtttgg tgggtgaatt ggcgaaggga ttgaaggagg catgggact 780
 taaccttttt aacgttgatg tcattagaga tggtaaggaa ccgacaaggt acctcgttat 840
 tgatatacat tactttcccg ggtatgcgaa attgccctct tatgagcctt ttatcacoga 900
 ttttttgttg gacattgtac gctccaagac tgcgtagtgc gtaggcaaat agttgaaagt 960
 ccgaggatga tcgatgatgt tattaccacat ttacactggt gtctaatcat tttgtttatg 1020
 ctgttttaaa aaaaaaaaaa
 <210> 22
 <211> 313
 <212> PRT
 <213> Glycine max
 <400> 22
 Ile Arg His Gln Ala Gly Gln Arg Tyr Arg Val Gly Tyr Ala Leu Gln
 1 5 10 15
 Gly Lys Lys Val Glu Ser Phe Ile Gln Pro Ser Leu Leu Asp His Ala
 20 25 30
 Lys Gln His Ser Ile Asp Leu Val Gln Ile Asp Pro Thr Ala Pro Leu
 35 40 45
 Gln Gln Gln Gly Pro Phe His Cys Ile Ile His Lys Leu His Thr Gln
 50 55 60
 His Trp Lys Asn Leu Leu Gln Gln Phe Ser Ser Lys His Pro Asn Thr
 65 70 75 80

Val Ile Ile Asp Pro Pro Glu Leu Val Asp Arg Leu His Asn Arg Val
85 90 95

Ser Met Leu Asp Ala Val Thr His Leu Gln Phe Ser Leu Glu Asn Ala
100 105 110

Thr Ile Gly Val Pro Lys Gln Val Val Val Asn Glu Pro Lys Ser Phe
115 120 125

Asp Leu His Lys Phe Glu Glu Gln Gly Leu Arg Phe Pro Val Ile
130 135 140

Ala Lys Pro Leu Ala Ala Asp Gly Gly Ala Gly Ser His Glu Leu Cys
145 150 155 160

Leu Val Phe Asp Glu Glu Gly Leu His Ala Leu Ser Val Pro Met Val
165 170 175

Leu Gln Glu Phe Val Asn His Gly Gly Val Val Phe Lys Ile Tyr Val
180 185 190

Ala Gly Gln Arg Val Asn Cys Val Lys Arg Lys Ser Leu Gly Asp Ile
195 200 205

Thr Glu Glu Lys Leu Lys Val Leu Arg Gly Ser Leu Pro Phe Ser Arg
210 215 220

Val Ser Ser Leu Gly Val Glu Asp Glu Gly Gly Gly Ala Val Glu Asp
225 230 235 240

Ala Glu Met Pro Pro Gln Ser Leu Val Gly Glu Leu Ala Arg Gly Leu
245 250 255

Arg Glu Ala Leu Gly Leu Asn Leu Phe Asn Val Asp Val Ile Arg Asp
260 265 270

Gly Lys Glu Pro Thr Arg Tyr Leu Val Ile Asp Ile Asn Tyr Phe Pro
275 280 285

Gly Tyr Ala Lys Leu Pro Ser Tyr Glu Pro Phe Ile Thr Asp Phe Leu
290 295 300

Leu Asp Ile Val Arg Ser Lys Thr Ala
305 310

<210> 23
<211> 522
<212> DNA
<213> Triticum aestivum

<400> 23
gcacgaggcc aagcagagtc gagcagcgca tcaaatcatg gtggccgacc acccagtcct 60
ccccgcggcc ggggtcacc atcggtctacg cgtgcgcgcc cggcaaggcg ggcagcgta 120
tccagcggcc gcttgaggcc ctgcggcgcg agcgcggcat gcgcctcgtc gccgtcgagc 180
cctcgctgcc cctggccgac cagggccctc tcgacctcat catccacaag ctcttcgacc 240
ggccctggcg gcgcgagctg gaggcctctt ccgcgctcca cccctccgtg cccgtcgctg 300
acgccccgc gcgcgtcgac cgctgctcg accgcttcac catgctcgac gtcgtccccg 360
ggctgcggcg cgccctggac ttcccgctca cgcgtccccc gcaagtcacc gtgaagcgac 420
ggcgcggcg tgcccgcgga cgaccgctcc caggggtccc gctcccgctc atcgccaagg 480
cgtcgccgtg cgacggagcg caactccaag actggctgtg tt 522

<210> 24
<211> 160
<212> PRT
<213> Triticum aestivum

<400> 24
Met Val Ala Asp His Gln Ser Ser Pro Arg Pro Arg Val Thr Ile Gly
1 5 10 15
Tyr Ala Leu Pro Pro Gly Lys Ala Gly Ser Val Ile Gln Pro Pro Leu
20 25 30
Glu Ala Leu Ala Ala Glu Arg Gly Met Arg Leu Val Ala Val Asp Ala
35 40 45
Ser Leu Pro Leu Ala Asp Gln Gly Pro Phe Asp Leu Ile Ile His Lys
50 55 60
Leu Phe Asp Arg Pro Trp Arg Ala Gln Leu Glu Ala Phe Ser Ala Leu
65 70 75 80
His Pro Ser Val Pro Val Val Asp Ala Pro Ala Ala Val Asp Arg Leu
85 90 95
Leu Asp Arg Phe Thr Met Leu Asp Val Val Pro Gly Leu Ala Ala Gly
100 105 110
Leu Asp Phe Pro Leu Ser Val Pro Ala Gln Val Thr Val Lys Arg Arg
115 120 125
Arg Arg Ala Gly Arg Gly Arg Pro Ser Thr Gly Ser Leu Pro Leu Ile
130 135 140
Ala Lys Pro Leu Ala Ser Thr Glu Arg Asn Ser Lys Thr Gly Cys Val
145 150 155 160

<210> 25
<211> 511
<212> DNA
<213> Triticum aestivum

<400> 25
cggccgcgcc gtctgcgtcc gccgcagcag ctgcccggac gtgcccgcgg aggcgcctcg 60
cggaacccga cgcggagcgc tcggtccccc tcgccaacat ctccagccgc cccgcgctcg 120
acaaggggga ggaactcgatg ccgcccgcgg cgttcgtgga ccagggtggc cgcgggctcc 180
ggcaggcgct ggggctgcac ctccctaact tcgacatgtt cgcggcgacg gagctggagc 240
acggcgccgg cgggaggtac ttccctcgtag acatcaacta ctcccggggg ttccgccaaga 300
tgccggcgcta cgagactgct ctcacgagatt tcttcgcccga gatgattcag ctaggcactg 360
caggcgccgc cgccggccaa gagaagctcg aactctgtgc ctgcgaatgag ctctaggaaa 420
ctgggtttac ttgggacttt tgatgaacag ggtgtcccta ctgccgagaa taaaatttaa 480
gaacgctttg cggcgtaaaa aaaaaaaaaa a 511

210> 26
<211> 105
<212> PRT
<213> Triticum aestivum

<400> 26
Ile Ser Ser Arg Pro Ala Leu Asp Lys Gly Glu Asp Ser Met Pro Pro
1 5 10 15
Ala Ala Phe Val Asp Gln Val Ala Arg Gly Leu Arg Gln Ala Leu Gly
20 25 30
Leu His Leu Leu Asn Phe Asp Met Phe Ala Ala Thr Glu Leu Asp Asp
35 40 45
Gly Gly Arg Arg Arg Tyr Phe Leu Val Asp Ile Asn Tyr Phe Pro Gly
50 55 60

Phe Ala Lys Met Pro Gly Tyr Glu Thr Ala Leu Thr Asp Phe Phe Ala
65 70 75 80

Glu Met Ile Gln Leu Gly Thr Ala Gly Ala Ala Gly Gln Glu Lys
85 90 95

Leu Glu Ser Val Pro Cys Asn Glu Leu
100 105

<210> 27
<211> 505
<212> DNA
<213> Triticum aestivum

<400> 27
gcacgaggat caaggccctt ttgatgttat tctgcataag ttgactggaa aggagtggca 60
acgacgggtg gaggaatata gagacacaca cccagaagtt actgttcttg acccaccagg 120
tgccatagaa catttgctga accgccaatc tatgcttcaa gaagtttcta aactggacct 180
tcgggattgt catggtaaag ttggtgttcc gaagcagctt ttgttaaca cagacccatt 240
gtcaatacca gctgcggtta tgagggtggt gctatcgctt ccattagtgg cgaagccctt 300
ggtggcgaag tctcatgagt tgcctcttgc ttatgactca gcttcccttga cgaagcttga 360
gcctccgttg gttcttcagg aatttgtaa ccatggcggc gctctattta aggtctacat 420
tgttggggat gcaattaggg tagtgcgtag gttttcactg cctaacgtgg atgacggtga 480
cctgtcaaat aacgcccagg tgttc 505

<210> 28
<211> 159
<212> PRT
<213> Triticum aestivum

<400> 28
His Glu Asp Gln Gly Pro Phe Asp Val Ile Leu His Lys Leu Thr Gly
1 5 10 15

Lys Glu Trp Gln Arg Arg Leu Glu Gly Tyr Arg Asp Thr His Pro Glu
20 25 30

Val Thr Val Leu Asp Pro Pro Gly Ala Ile Glu His Leu Leu Asn Arg
35 40 45

Gln Ser Met Leu Gln Glu Val Ser Lys Leu Asp Leu Ala Asp Cys His
50 55 60

Gly Lys Val Gly Val Pro Lys Gln Leu Phe Val Asn Thr Asp Pro Leu
65 70 75 80

Ser Ile Pro Ala Ala Val Met Arg Ala Gly Leu Ser Leu Pro Leu Val
85 90 95

Ala Lys Pro Leu Val Ala Lys Ser His Glu Leu Ser Leu Ala Tyr Asp
100 105 110

Ser Ala Ser Leu Thr Lys Leu Glu Pro Pro Leu Val Leu Gln Glu Phe
115 120 125

Val Asn His Gly Gly Val Leu Phe Lys Val Tyr Ile Val Gly Asp Ala
130 135 140

Ile Arg Val Val Arg Arg Phe Ser Leu Pro Asn Val Asp Asp Gly
145 150 155

<210> 29
<211> 638
<212> DNA
<213> Triticum aestivum


```

<400> 29
ccggatcagg tagtggcaac attggetgcg cgcgcgactg ggtcttagac tgttcatcat 60
acctatgac agagagcatg ggacacggga tcggttttat gttatagaca tgaactattt 120
tcctgggtat gggaagatgc ctggatacga gcatgttttc accgacttcc tgctgagcct 180
agaccagcag aaggagtaca agcgacgact gggctatagc tcgggtgagg ggtgaagagt 240
cgaggecgac caacttccct ctgacctttg ccagtattgt gtctcttcgc attcgctgtg 300
tgtacgaagc aatgtgcata tcaggagggg aaggatcatt aagtttcatc tgcgttagag 360
ttgggcatca tagtttgca tgaataaggca gcgaggactg tgttgaactg taattattaa 420
tcgcattttt ggggcaactg tgtttcttct atgtctcttga ataactctctg tcggcttagc 480
aatggaatcg agcgggtatc cgcacccaca tcagggtcttc cagaaatctc aaagtattgt 540
agacaatccc agcgaatatt tccctgctga cgttggtgtg tgcactggcg gaggtcttgg 600
ctaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaa

```

```

<210> 30
<211> 67
<212> PRT
<213> Triticum aestivum

```

```

<400> 30
Arg Arg Leu Gly Leu Arg Leu Phe Ile Ile Ser Met Ile Arg Glu His
1 5 10 15
Gly Thr Arg Asp Arg Phe Tyr Val Ile Asp Met Asn Tyr Phe Pro Gly
20 25 30
Tyr Gly Lys Met Pro Gly Tyr Glu His Val Phe Thr Asp Phe Leu Leu
35 40 45
Ser Leu Asp Gln Gln Lys Glu Tyr Lys Arg Arg Leu Gly Tyr Thr Ser
50 55 60

```

```

Gly Glu Gly
65

```

```

<210> 31
<211> 1141
<212> DNA
<213> Triticum aestivum

```

```

<400> 31
tttcatacca attgatgaga ctgcctctct ctcgaacaaa ggccatttgg atattatctt 60
gcacaaagaaa actagcaagg agtggcagcg gtttctggag gattatcatg aagtgcattc 120
agaagtcact gtctctgacc caccacaatgc cattgagcat ctgaacaatc gacaatcgat 180
gcttgaagaa gtacagagatt tgaacctgtc cagtttctat gaagaagttt gcaactccag 240
ccaactgtgc attatgaag atccgttctc tataccaact gcagtgtcca tggctgggt 300
aaacttgcca ttgggtgcca agccalttgg tgttgatgga acatctaaat ctcatgagct 360
atctctggcca taagatgagg cgtcttctaa ggtgtacatc attggtgaag ctatacaggt 420
atttgtgaat catggggggg atctctttaa acccttggcc ttactaaaca attgttgcat 480
tgccgcagg ttctctcttc ctgatgttaa accttggcc tatctaaaca attgttgcat 540
ctttggactt ccagggggtt ctgtcggttc agcaagtctc gatcatgacg cccctgatcc 600
tcgtattgca gagcttcccc cgagaccact tctagagaaa ctaggcaggg agcgtcgttg 660
tgcgttgggt ctaagattgt tcaacataga tatgattaga gagcttggag caaatgaccg 720
gtactatata attgatatca actacttccc aggatattgg aaaatgcctg gctatgagca 780
catattcacc gatctcttgc aaagtcttgg gcaaaaacaa taccaaaggt gcttgagcgg 840
aggctgaagc gctggtagcc atacttgttc actaatgtgc atactttgga tacataagct 900
ggagcccagg cgttgaagtc ccaaaacttac atgttttgtt agaaatggcg ccaatcaccc 960
agggtcgagg gtaggggagg agaaagaaga actcgaatgg ccttctgccc atctgttata 1020
aatcacgggt gttttgatca tagtttggtc tcaaatgctt tcatcaaca ggaatcagac 1080
catcaaccca ccatgtatgt ggctcgaggg gggggcccta cacaatcaaa ccttgggggc 1140
c 1141

```

```

<210> 32
<211> 281
<212> PRT
<213> Triticum aestivum

```

Arg Gln Arg Gly Ile Asp Leu Val Lys Leu Asp Pro Thr Lys Ser Leu
 35 40 45
 Leu Glu Gln Gly Lys Leu Asp Cys Ile Ile His Lys Leu Tyr Asp Val
 50 55 60
 Tyr Trp Lys Glu Asn Leu His Glu Phe Arg Glu Lys Cys Pro Gly Val
 65 70 75 80
 Pro Val Ile Asp Leu Pro Glu Ala Ile Glu Arg Leu His Asn Arg Val
 85 90 95
 Ser Met Leu Glu Val Ile Thr Gln Leu Arg Phe Pro Val Ser Asp Ser
 100 105 110
 Glu Arg Phe Gly Val Pro Glu Gln Val Val Met Asp Ser Ser Val
 115 120 125
 Leu Ser Gly Gly Gly Ala Leu Gly Glu Leu Lys Phe Pro Val Ile Ala
 130 135 140
 Lys Pro Leu Asp Ala Asp Gly Ser Ala Lys Ser His Lys Met Phe Leu
 145 150 155 160
 Ile Tyr Asp Gln Glu Gly Met Lys Ile Leu Lys Ala Pro Ile Val Leu
 165 170 175
 Gln Glu Phe Val Asn His Gly Gly Val Ile Phe Lys Val Tyr Val Val
 180 185 190
 Gly Asp His Val Gln Cys Val Lys Arg Arg Ser Leu Pro Asp Ile Ser
 195 200 205
 Glu Glu Lys Ile Gly Thr Ser Lys Gly Ser Leu Pro Phe Ser Gln Ile
 210 215 220
 Ser Asn Leu Thr Ala Gln Glu Asp Lys Asn Ile Glu Tyr Gly Glu Asp
 225 230 235 240
 Arg Ser Leu Glu Lys Val Glu Met Pro Pro Leu Ser Phe Leu Thr Asp
 245 250 255
 Leu Ala Lys Ala Met Arg Glu Ser Met Gly Leu Asn Leu Phe Asn Phe
 260 265 270
 Asp Val Ile Arg Asp Ala Lys Asp Ala Asn Arg Tyr Leu Ile Ile Asp
 275 280 285
 Ile Asn Tyr Phe Pro Gly Tyr Ala Lys Met Pro Ser Tyr Glu Pro Val
 290 295 300
 Leu Thr Glu Phe Phe Trp Asp Met Val Thr Lys Lys Asn His Val
 305 310 315
 <210> 34
 <211> 319
 <212> PRT
 <213> Arabidopsis thaliana
 <400> 34
 Met Ser Asp Ser Ile Gln Glu Arg Tyr Leu Val Gly Tyr Ala Leu Ala
 1 5 10 15
 Ala Lys Lys Gln His Ser Phe Ile Gln Pro Ser Leu Ile Glu His Ser
 20 25 30

Arg Gln Arg Gly Ile Asp Leu Val Lys Leu Asp Pro Thr Lys Ser Leu
 35 40 45
 Leu Glu Gln Gly Lys Leu Asp Cys Ile Ile His Lys Leu Tyr Asp Val
 50 55 60
 Tyr Trp Lys Glu Asn Leu His Glu Phe Arg Glu Lys Cys Pro Gly Val
 65 70 75 80
 Pro Val Ile Asp Leu Pro Glu Ala Ile Glu Arg Leu His Asn Arg Val
 85 90 95
 Ser Met Leu Glu Val Ile Thr Gln Leu Arg Phe Pro Val Ser Asp Ser
 100 105 110
 Glu Arg Phe Gly Val Pro Glu Gln Val Val Val Met Asp Ser Ser Val
 115 120 125
 Leu Ser Gly Gly Gly Ala Leu Gly Glu Leu Lys Phe Pro Val Ile Ala
 130 135 140
 Lys Pro Leu Asp Ala Asp Gly Ser Ala Lys Ser His Lys Met Phe Leu
 145 150 155 160
 Ile Tyr Asp Gln Gln Gly Met Lys Ile Leu Lys Ala Pro Ile Val Leu
 165 170 175
 Gln Glu Phe Val Asn His Gly Gly Val Ile Phe Lys Val Tyr Val Val
 180 185 190
 Gly Asp His Val Lys Cys Val Lys Arg Arg Ser Leu Pro Asp Ile Ser
 195 200 205
 Glu Glu Lys Ile Gly Thr Ser Lys Gly Ser Leu Pro Phe Ser Gln Ile
 210 215 220
 Ser Asn Leu Thr Ala Gln Glu Asp Lys Asn Ile Glu Tyr Gly Glu Asp
 225 230 235 240
 Arg Ser Leu Glu Lys Val Glu Met Pro Pro Leu Ser Phe Leu Thr Asp
 245 250 255
 Leu Ala Lys Ala Met Arg Glu Ser Met Gly Leu Asn Leu Phe Asn Phe
 260 265 270
 Asp Val Ile Arg Asp Ala Lys Asp Ala Asn Arg Tyr Leu Ile Ile Asp
 275 280 285
 Ile Asn Tyr Phe Pro Gly Tyr Ala Lys Met Pro Ser Tyr Glu Pro Val
 290 295 300
 Leu Thr Glu Phe Phe Trp Asp Met Val Thr Lys Lys Asn His Val
 305 310 315